V-CoHOG: Volumetric Co-occurrence Histograms of Oriented Gradients for Texture Classification in Medical Imaging

by

Alexander Filbert

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science
University of Alberta

Abstract

Image textures, which are properties that can describe the pixel intensities of an image, have been analyzed in order to perform a variety of tasks, including segmentation and classification. Features derived from textures which are invariant to changes such as contrast and transformations are of particular interest, due to their comparability across different images with related content. Hence, even early methods such as Gray-Level Co-occurrence Matrices (GLCM) have sought to describe textures using invariant features which proved to be effective in classification experiments. With the development of gradient based methods such as Histograms of Oriented Gradients (HOG) and derivatives including Co-occurrence Histograms of Oriented Gradients (CoHOG), an increasing variety of texture analysis methods have been applied to various domains, including medical imaging. When tasked with the classification of medical patients and healthy controls, methods such as Modified Co-occurrence Histograms of Oriented Gradients (M-CoHOG), which was applied to datasets of the neurodegenerative disorder Amyotrophic Lateral Sclerosis (ALS), have attained leading performance metrics. However, in the case of ALS, which can be difficult to diagnose, the lack of known biological imaging markers which could streamline the selection process of a region of interest (ROI), necessitates the contribution of an expert to segment the region of interest for M-CoHOG. Additionally, the consideration of only 2D imaging slices results in features with decreased spatial descriptiveness compared to volumetric approaches. In this thesis, a volumetric feature extraction method called V-CoHOG, extended on M-CoHOG, with automated ROI extraction is proposed. To reduce the feature vector size resulting from V-CoHOG, a feature

selection method based on ReliefF is applied before classification using an ensemble model classifier with 5 base classifiers. The Canadian ALS Neuroimaging Consortium 1 (CALSNIC-1) and, for the first time, CALSNIC-2 ALS datasets are used for evaluation and comparison with four 3-dimensional convolutional neural network (CNN) methods, with the additional Alzheimer's Disease Neuroimaging Initiative (ADNI) Alzheimer's disease dataset used to demonstrate the versatility of the proposed method. Analysis of the results demonstrates that the proposed method is able to consistently outperform the 3D CNN approaches on ALS datasets, while approaching and sometimes outperforming M-CoHOG performance on the CALSNIC-1 datasets. Furthermore, a CUDA-accelerated implementation of M-CoHOG with significantly improved runtime performance on higher-resolution images is proposed. Finally, segmentation maps of selected features for ALS classification are overlaid onto the original imaging volumes. These maps allowed for the investigation of potential ALS imaging biomarkers, specifically their localization to brain regions present in the ROI.

Acknowledgments

I would like to extend my sincerest gratitude to my supervisors, Dr. Herbert Yang, Professor, Department of Computing Science, and Dr. Sanjay Kalra, Professor, Department of Medicine, Biomedical Engineering, and Computing Science, whose support was crucial throughout the development of this thesis. Their encouragement and openness to dialogue allowed me to continually elevate the scope and granularity of my research. I can speak only fondly of the time I spent with them.

I would like to thank Dr. Pierre Boulanger, Professor, Department of Computing Science, for his willingness to serve on my examining committee.

I had the pleasure of working with the members of the Computer Graphics and ALSNIRU Labs, and I want to thank them all for their support. In particular, I want to thank Rafsanjany Kushol and G.M. Mashrur E Elahi for their advice.

I would also like to thank the ALS Society of Canada, Brain Canada, and NSERC for their generous financial support.

I am grateful to my family and friends, whose encouragement facilitated the research that was essential for this thesis.

Finally, and most importantly, I would like to thank God for answering my prayers and giving me the chance to be involved in this work. Without him, none of this would have been possible.

Table of Contents

List of Tables

List of Figures

Abbreviations & Acronyms

1D One-Dimensional

2D Two-Dimensional

3D Three-Dimensional

AD Alzheimer's Disease

ADNI Alzheimer's Disease Neuroimaging Initiative

ALS Amyotrophic Lateral Sclerosis

BET Brain Extraction Tool

CALSNIC Canadian ALS Neuroimaging Consortium

CM Co-occurrence Matrix

CNS Central Nervous System

CoHOG Co-occurrence Histograms of Oriented Gradients

CPU Central Processing Unit

CST Corticospinal Tract

DTI Diffusion Tensor Imaging

FLAIR Fluid-Attenuated Inversion Recovery

GL Gray-Level

GLCM Gray-Level Co-occurrence Matrix

GO Gradient Orientation

GPU Graphics Processing Unit

HOG Histograms of Oriented Gradients

LMN Lower Motor Neuron

M-CoHOG Modified Co-occurrence Histograms of Oriented Gradients

MRI Magnetic Resonance Imaging

MS Multiple Sclerosis

PNS Peripheral Nervous System

ROI Region Of Interest

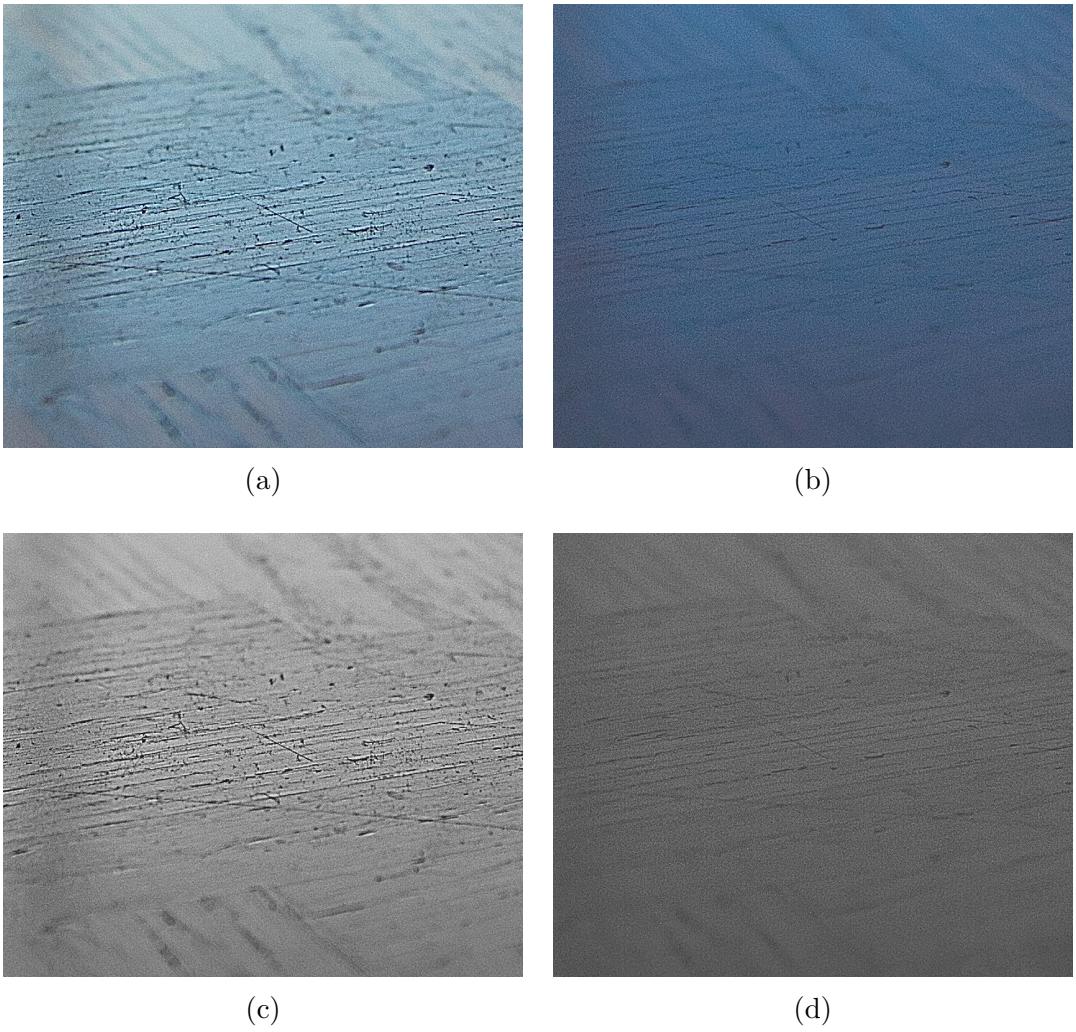
UMN Upper Motor Neuron

V-CoHOG Volumetric Co-occurrence Histograms of Oriented Gradients

Chapter 1

Introduction

As one of the five senses, vision has an irreplaceable role in allowing humans to perceive their environment and to effectively interact with it. It is no surprise, therefore, that many early approaches to add visual recognition capabilities to computers and automated systems sought to mimic various aspects of human eyesight, potentially enabling them to compute statistics or act on such information in a desired manner. Because modern computers are digital, computer vision tasks normally rely on digital images, consisting of pixel matrices, as input information. There are various properties in digital images which a human may quantify in order to perform a recognition task, including colour, shape and brightness. Another such property is textures, which vary from fine, when considering smooth surfaces, to coarse, like a rough stone wall. Further texture properties, such as their spatial appearance, allow for classification as structured or stochastic types: A brick wall with a regular pattern displays a clear structure in contrast to the stochastic and uneven nature of a marble design. In radar imagery, fine texture can be indicative of fine-grained sedimentary rocks [1] allowing for aerial surveying. This is but one example of how different textures can provide valuable insight into objects captured in a photograph.



(a) (b)

(c) (d)

Figure 1.1: Photograph of a wooden table under two different lighting conditions. The top row presents the original colour image while the bottom row has been desaturated such that the gray levels are more apparent. The texture of the surface remains consistent across all images. Photograph courtesy of Evan Martens.

1.1 Image Textures

In digital images, textures can be expressed as the statistical distribution of gray levels, or intensities, in an image [1]. Hence, textures and the intensities of image pixels cannot be separated when considering image properties. Nevertheless, changes in intensities do not indicate a change in textures: For example, two photographs taken of the same wooden table inside a room with a light turned on and off, respectively. When the light is turned on, the image pixels will have a greater intensity, or

brightness, than when the light is turned off. This effect will also be present in the pixels corresponding to the table in the image. However, the texture of the wood will remain the same between the images (Figure 1.1). This phenomenon illustrates why textures have a significant potential as an informative property in images, even when other characteristics, such as the brightness, colour or even orientation change.

1.2 Texture Features

Computer vision relies on such properties, referred to as features, to perform tasks such as object recognition [1] and image classification [2]. A variety of approaches have been developed to analyze texture in images and it is invariant methods in particular that appear to show the most promise as features for downstream tasks. Invariant texture analysis methods can generally be categorized as either structural, model, or statistical [3], however, they all share a resistance to translations, rotations, as well as affine and perspective transformations. This attribute is essential when considering texture features for classification tasks where images of a target class or object may differ due to the aforementioned image transformations. Statistical methods describe texture via statistics of selected features. Examples include a variety of high order statistics as well as polar plots and harmonic expansion [4]. Model methods attempt to describe texture in terms of probability models or linear combinations of basis function sets and include wavelet transform operations [5] and Markov models [6]. Lastly, structure methods arrange texture elements according to placement rules and include such methods as invariant histograms [3] and morphological decomposition [7].

1.3 Motivation

Given the presence of image textures in digital images with extensively different characteristics, texture analysis has seen applications in a variety of domains, with

tasks such as object recognition and image classification. In medical imaging, texture analysis has allowed for advances in the classification of subjects with conditions such as Alzheimer's Disease (AD) and Amyotrophic Lateral Sclerosis (ALS) which affect the brain and can be difficult to diagnose. Modified Co-occurrence Histograms of Oriented Gradients (M-CoHOG) [8], a method that relies on texture analysis to generate classification features of ALS patients, has previously allowed for state-of-the-art classification performance of the CALSNIC-1 ALS dataset. However, M-CoHOG can only be applied to Magnetic Resonance Imaging (MRI) image slices, even though MRI systems generate volumetric imaging data. This limitation requires an expert to manually select and segment image slices for processing, significantly increasing the method's data preparation requirements. Hence, an approach which could match or exceed M-CoHOG's classification performance without requiring the contributions of an expert could prove to be a substantial advancement.

1.4 Contribution Summary

In the process of attempting to address the limitations of the M-CoHOG method, a number of developments and observations were made with the main contributions of this thesis being:

- GPU Accelerated M-CoHOG: The M-CoHOG algorithm was examined and the two costliest, in terms of runtime, components were identified to be the determination of the gradient orientation as well as the co-occurrence matrix calculation. Hence, this thesis introduces a graphics processing unit (GPU) implementation on the Nvidia CUDA platform for these sections of the M-CoHOG, which was able to significantly speed-up the feature generation process on high resolution images. Due to the steady resolution increases of image acquisition devices such as digital cameras or medical scanners, an accelerated approach as is proposed is expected to become increasingly necessary for this

type of gradient based feature extraction.

- Classification of ALS using HOG-derived features on ALS 3D volumetric data: This thesis proposes a gradient-based texture feature extraction method that is able operate on volumetric imaging data in order to extract features based on Histograms of Oriented Gradients (HOG). Specifically, the proposed approach, Volumetric Co-occurrence Histograms of Oriented Gradients (V-CoHOG) develops the concepts introduced by Co-occurrence Histograms of Oriented Gradients (CoHOG) and expanded upon via M-CoHOG into a method capable of generating features for volumes derived from MRI systems. After a reduced feature set is selected and used for the evaluation with the proposed ensemble classifier model, superior classification performance compared to 4 3D Convolutional Neural Network (CNN) approaches can be achieved.

- Automation of ROI selection based on segmentation masks generated by Brain Extraction Tool (BET): Rather than rely on registration and normalization, which can cause significant changes to the original imaging data, or individual segmentations by an expert, which can be laborious and time-consuming, this thesis proposes an automated ROI selection process that does not alter the imaging data beyond the utilization of a cubic scaling factor. The Brain Extraction Tool (BET) is used to generate brain segmentation masks which are then used in conjunction with offset parameters in order to facilitate comparable ROI cropping across samples. Furthermore, the experimental results demonstrate that the proposed method is able to approach and even surpass, depending on the data subset, the classification results achieved by M-CoHOG using manually selected ROIs.

- First application of the CALSNIC-2 dataset in a classification study: While the multi-centre CALSNIC-1 dataset has been used for ALS classification experiments before, the presently larger CALSNIC-2 dataset has not yet been subject

of such studies. The classification results attained on the CALSNIC-2 dataset using the method proposed by this thesis can, therefore, act as a baseline for studies going forward. We expect future research into ALS to make use of continually larger datasets, which is why the CALSNIC-2 results presented could act as a baseline for future classification approaches.

- Feature segmentation maps detailing the region of interest of selected features: Using the optimal subset of features selected for classification, segmentation maps of the corresponding voxels in the computed gradient orientation (GO) are scaled up and overlayed onto the input image volume. The resulting heatmaps, which indicate regions acting as the origin of multiple features with an increased intensity, provide insight into the discernability, via texture features, of ALS patient and controls allowed for by these regions. Hence, the feature segmentation results presented in this thesis can act as a tool in the search for biological imaging markers of ALS, which could facilitate accelerated patient assessment and diagnosis.

1.5 Outline

The thesis is structured into 5 main chapters as follows:

In Chapter 2, we discuss a variety of literature in order to establish premises regarding methods for texture analysis, accelerated image processing and Magnetic Resonance Imaging (MRI) with a focus on Amyotrophic Lateral Sclerosis (ALS).

Chapter 3 discusses the methods applied in the experiments conducted as part of this thesis, including parallel algorithm implementations, feature extraction, feature selection and classification.

In Chapter 4, we present and discuss our results and how they compare to alternative methods.

Finally, in Chapter 5, a summary of the significant results is provided along with

the conclusion of the thesis.

Chapter 2

Background

2.1 Invariant Texture Features

2.1.1 Gray Level Co-occurrence Matrix

Not long after the advent of digital imaging, Haralick discovered that a number of statistical texture features could be derived from 4 angular nearest-neighbour gray-tone spatial-dependence matrices computed from an image [1]. Since then, these matrices have come to be known as the Gray-Level Co-occurrence Matrices (GLCM), with the term "gray-level" being used to describe pixel intensities, while the features themselves are now often referred to as Haralick features. Using resolution cells, of which each non-edge nearest-neighbour cell has 8 neighbours, a total of 28 textural features were suggested. Of particular interest are features such as entropy and maximal-correlation which proved to be invariant under monotonic gray-tone transformations. Hence, GLCM features involve counting the number of pixel pairs that share certain relationships or values in order to create a matrix that is able to represent an image's texture. The specifics of the pixel pairs depend on the feature, for example, entropy f is calculated as

$$f = -\sum_{i}\sum_{j} p(i,j)\log(p(i,j)), \tag{2.1}$$

where $p(i,j)$ is the (i,j)th entry in the normalized co-occurrence matrix.

2.1.2 Gray Level Run Lengths

Expanding on the Haralick texture features computed using gray-levels, a related set of features calculated using Gray-Level Run Lengths (GLRL) were developed [1]. Gray-Level Runs refer to collinear picture points that share a gray level value and are located in succession. A matrix, computed using GLRL for a particular direction, can then be used with 5 proposed functions in order to compute a set of measures useful for classification. Specifically, the 5 functions suggested are Short Runs Emphasis, Log Runs Emphasis, Gray Level Nonuniformity, Run Length Nonuniformity, and Run Percentage. The Run Percentage refers to the ratio of total runs to total possible runs, assuming every run has length equal to 1.

2.1.3 Wavelets and Scale Invariant Feature Transform

While texture features based on GLCM and GLRL continue to be applied, newer methods have given rise to descriptors such as wavelets [9] and Scale Invariant Feature Transform (SIFT) [10]. Using a decomposition in order to transform a 1-D signal onto a basis of wavelet functions, wavelet image analysis allows for the calculation of features that have found numerous applications in the study of medical images, particularly in segmentation i.e. sectioning tasks [11]. Instead, SIFT descriptors use operations to assign location, scale and orientation measurements to keypoints derived from an image. These measurements are then used in tandem to create a 2D coordinate system to describe a local image region that is invariant to location, scale and orientation parameters.

2.1.4 Histograms of Oriented Gradients

Histograms of Oriented Gradients (HOG) is a result of further developments in texture analysis based on using gradient orientations [12]. A locally normalized implementation allows for clear classification performance advantages over approaches that use feature sets such as wavelets. The traditional HOG method involves a localized

approach in which an image, whose gradient orientations have been quantized into N bins, is divided into cells with a histogram computed for each cell. Calculation of the histogram requires only the tallying of different orientation groups within the cell. Finally, to obtain the feature set representing the whole picture or region of interest (ROI), the computed histograms are concatenated into a single vector. To counterbalance the localized nature of the approach, rectangular (R-HOG) and circular (C-HOG) descriptor blocks consisting of grouped cells are subjected to local contrast normalization. An overlap in the blocks also increases performance with the circular descriptors having a slight edge.

Table 2.1: Parameters used for the circular descriptor blocks in C-HOG [12].

C-HOG Layout Parameter	Default
Number of Angular Bins	At least 4
Number of Radial Bins	At least 2
Radius of Central Bin (pixels)	-
Expansion Factor for Subsequent Radii	-

2.1.5 Co-occurrence Histograms of Oriented Gradients

As a way to further enhance the generalizability of HOG features across image regions, Co-occurrence Histograms of Oriented Gradients (CoHOG), which involve histograms computed using gradient orientation pairs, were proposed [13]. Specifically, a co-occurrence matrix (CM) is used to describe the distribution of gradient orientations at a particular offset, creating a sliding-window algorithm. Hence, as with classic HOG, the gradient orientations of the image must initially be determined, after which the image can be split into cells. Watanabe et al. used both the Sobel and Roberts filter for this;

$$\Theta = \arctan\left(\frac{G_v}{G_h}\right), \tag{2.2}$$

where Θ is the gradient orientation, G_v the vertical gradient and G_h the horizontal gradient. Each pixel is then labelled as having no gradient, as it was below a certain threshold, or belonging to one of 8 orientation bins. To cover a total of 360 degrees, each orientation bin spans 45 degrees. The co-occurrence matrices for every cell are determined by considering all pixels in the cell as well as the number of offsets dictating the number of orientation pairs (Figure 2.1). The number of offsets, which can be described as the neighbourhood size in terms of radius R resulting in a particular number of pixels B surrounding a pixel of interest A, can describe shapes that are both more local or global relative to the current pixel A, depending on R. When applied to a pedestrian detection task, CoHOG performs superior to the original HOG method, even with human detection being the primary task used to test HOG, possibly because of the enhanced ability to capture more local and global information with offsets. Nevertheless, the method retains the use of cells which may still limit performance for cell boundary areas.

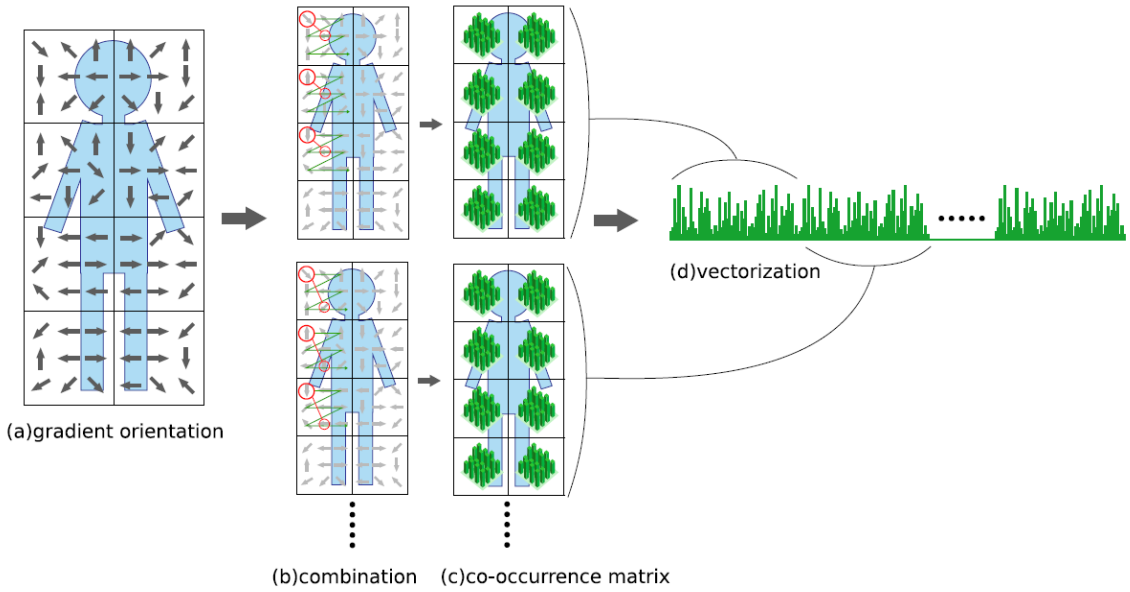


Figure 2.1: Generation of the feature vector in CoHOG [13]. After the gradient orientation is determined using the Sobel or Roberts filter, the co-occurrence matrices are computed by counting the number of orientation pairs and adding them to the respective bins. Finally, the CMs are vectorized and concatenated into the feature vector.

2.1.6 Other HOG Extensions

Although CoHOG presents an interesting development over the original HOG, a number of other modifications on the original method have seen various degrees of success in a variety of applications. These include examples such as a pedestrian detection method which uses HOG in quaternion space [14] and moving human detection in real-time using not only HOG but also Fourier descriptors [15]. Additionally, extensions of the CoHOG method have also been investigated: Weighted CoHOG adds magnitude values to the feature vector using a weight function for improved human detection [16] while Multiresolution CoHOG (MRCoHOG) uses multiple low-resolution versions of the same image to generate a feature vector [17]. Various adaptations of the HOG descriptor have also been applied to detect tremors in hand-written drawings of Parkinson's disease patients [2] as well as for the recognition of places with changing environments, including hindrances from varying viewpoints, conditions and low-informative scenes [18]. HOG features have also been extended to the third dimension [19] for application in a road user classification system where spatial modelling allowed for improved feature extraction. Therefore, it is apparent that variations on the HOG descriptor for texture features are broadly applicable to a wide range of image datasets from which texture information can be derived.

2.2 Hardware Accelerated Image Processing

Digital images are constructed of a matrix of pixels, with each pixel potentially having multiple associated numerical values according to the number of channels present in the image. While a gray-scale image may consist of only one channel with each pixel's value corresponding to the intensity, or brightness, a colour image typically has three channels with each pixel encompassing a separate red, green and blue (RGB) value for the RGB channels. Depending on the data type, digital images are often thousands if not millions of pixels (megapixels) in size. Additionally, the standard approach for

computational tasks is to perform them sequentially, referred to as serial processing. Hence, an unmodified texture analysis approach that is applied to every pixel in the image will need to be executed thousands or millions of times, potentially resulting in an operation that is rather expensive in terms of computational time. This also makes such methods rather unfeasible for applications where any feature descriptors need to be determined within milliseconds, as in video object tracking. This is because a 30 frame-per-second video will display a single image (frame) for approximately 33 milliseconds until the next frame in the sequence is presented. Hence, for a tracking method to keep pace with video playback and be considered real-time, it must perform all processing operations in real-time.

2.2.1 Parallelization Requirements

The differences between Graphics Processing Units (GPUs) and Central Processing Units (CPUs) prevent the execution of standard texture analysis methods on graphics cards without modifications. However, segments of various algorithms including those used for GLCM and HOG can be successfully accelerated using GPUs [20–22]. This is because these algorithms fulfill an important requirement of parallelization: The operations that should be performed simultaneously are not dependent on the result of one another. Operations like the histogram computation using the gradient orientations in HOG rely on a sliding-window style algorithm where the orientation values within the value do not depend on orientations outside the window. Hence, all windows can be executed in parallel without affecting each other [22]. As the requirements for parallel execution are met, the portions of the algorithm which should be executed in parallel using a GPU can then be implemented using the C language along with an interface such as Nvidia's Compute Unified Device Architecture (CUDA) to allow for execution on Nvidia GPUs.

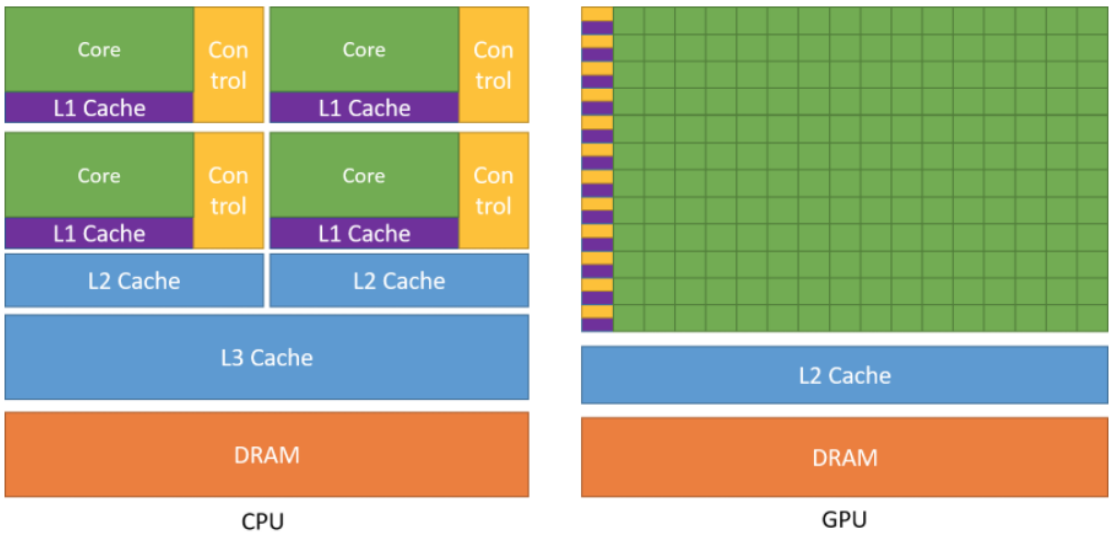


Figure 2.2: Abstract comparison of CPU and CPU architecture [23]. Note the significantly greater core count present on the GPU.

2.2.2 GPU Performance Benefits

While parallel implementations of texture analysis methods such as HOG are not strictly limited to GPUs, as is exemplified by a human detection system using an adaptation of CoHOG for field programmable gate arrays (FPGAs) [24], graphics cards are one of the most prevalent hardware components designed for some form of parallel computing. Hence, GPUs make an attractive execution platform for these accelerated approaches and have been used to successfully decrease the execution time requirements for a number of datasets. Execution speed improvements are highly dependent on various factors. However, a CUDA-accelerated version of HOG named fastHOG was able to increase execution speed by up to 67x on a single GPU [22], demonstrating the potential of such approaches. Particularly, the use of CUDA involves writing functions, referred to as kernels, which perform a certain task. A single kernel is mapped to a single execution thread on the GPU, with threads grouped into blocks and blocks grouped into grids [25]. In the case of fastHOG, a HOG pixel block was mapped to a CUDA thread block. Since modern GPUs may have thousands of CUDA cores, they can execute many threads at once, which also means that their

ability to access the memory which stores the information needed for processing will have big implications on the execution speed.

2.3 Magnetic Resonance Imaging

Until the development of computed tomography (CT) and magnetic resonance imaging (MRI) in the 1970s, no methods to properly scan and image the brain were available [26]. While CT makes use of x-rays for imaging, the MRI approach which relies on magnetic fields has shown to be superior for imaging many types of soft tissue [26]. More specifically, the MRI procedure begins with a superconducting magnet that creates a uniform magnetic field. Next, resistive electromagnets inside a bore are activated to create magnetic field gradients along three axes while coils, which transmit and receive radio waves, cause protons to resonate at certain magnetic field strengths. Subsequently, hydrogen nuclei positions along the axes are deduced by separating the radio signals using a Fourier transform. Hence, areas with more protons (hydrogen nuclei) cause an increase in signal amplitude, allowing for the gray levels in the image to indicate different signal intensities that correspond to tissue differences.

2.3.1 MRI Types

While the MRI scanner image acquisition technique has a degree of stability affected by parameters such as the magnetic field strength, the imaging type may have a drastic effect on the resultant picture data. There is a considerable difference between functional MRI (fMRI) and structural MRI (sMRI), for example, with functional MRI measuring intensity signals over time in order to record changes in blood flow, while sMRI attempts to capture tissue structure at a single point in time. sMRI can be further partitioned into types such as T1-weighted, T2-weighted and diffusion tensor imaging (DTI). T1 refers to the longitudinal relaxation time, or the time that it takes for spinning protons to realign with the scanner's external magnetic field [27]. In

contrast, T2 involves the transverse relaxation time; the time required for protons perpendicular to the magnetic field to lose phase coherence. The difference between these weightings allows for the visualization of tissue types with disparate contrast levels because of the water content nonuniformity in human tissues [28].

2.3.2 Post-processing of MRI Data

Since MRI images are constructed of signal intensity measurements, they typically consist of a single grayscale channel. As a way to accentuate areas of the image, gray-level (GL) transformations can be applied [29]. These transformations involve functions that alter the brightness (gray-level) of each pixel in an image. If a pixel A is brighter than another pixel B prior to a GL transformation, the relation will hold true after the function has been applied. However, the magnitude of difference, or contrast, may be affected. Therefore, GL transformation may increase contrast in certain image areas, but potentially at the expense of other regions. The intention of applying such operations to medical images is often to aid in the identification of biomarkers which are features of an image that are relevant to diagnosis. Specifically, contrast enhancement techniques such as S-curve gray-level transformation [30] can increase the perceptibility of edges between adjacent tissues, allowing for improved segmentation and classification performance in some cases.

2.3.3 MRI Dataset Limitations

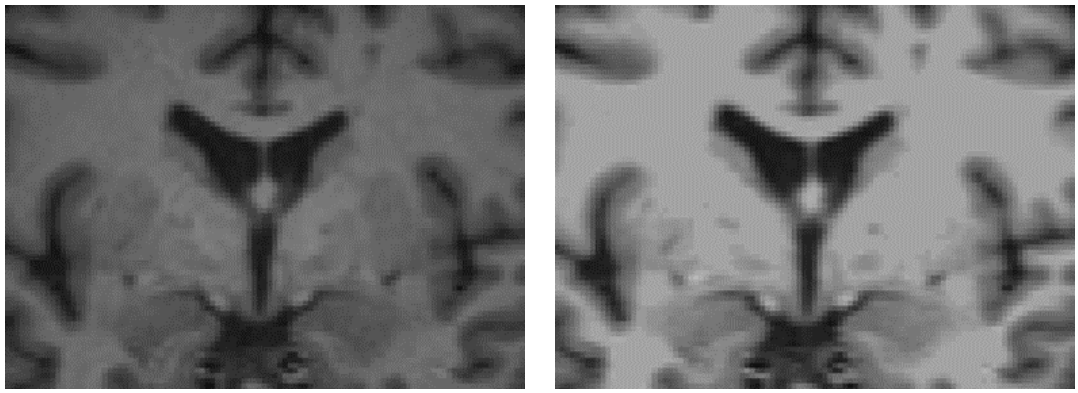
There are a multitude of conditions where image biomarkers can not be easily discerned by visual or computational inspection of raw and GL transformed imaging data. It is in such cases that texture analysis can provide valuable insight via the calculation of novel biomarkers that are applicable to patient diagnosis. Nevertheless, there is a considerable lack of standardization which is impacting scanner acquisition methodology involving differences in parameters as well as intrinsic scanner characteristics. Acquisition parameters can be standardized across systems, particularly

when data collection is coordinated. However, differences in scanner characteristics cannot be standardized. Even for a single scanner, the state of system components such as coil loading may change from one scanning session to the next. In T1 and T2-weighted imaging, for example, the signal intensity depends on factors such as the radio frequency and coil sensitivity that will all vary with the hardware used [31]. Additionally, while it may be possible to attain a certain degree of standardization with data acquired from a single imaging centre, the variance problems described are even more prevalent in datasets that attempt to combine images from multiple centres. Multicentre datasets are often the only way to achieve sample sizes larger than a few dozen scans for conditions such as amyotrophic lateral sclerosis (ALS). This highlights a dilemma present throughout the medical imaging domain where, for many cases, datasets containing thousands of samples do not exist. Attempts at data augmentation using synthetic data have been proposed to combat this deficiency [32, 33]. However, while generated data may be correlated with the other samples in the source dataset, this correlation may not accurately reflect the states of pathological biomarkers in the wild. Additionally, synthetic data may act as a source of data leakage if the derivation of the generated data is not restricted to the training set.

2.3.4 Data Normalization

Detailed differences in image characteristics such as voxel size are not always clear to the naked eye, but can have a significant impact on computed texture features including those derived from GLRLM and GLCM [34]. In many current models designed for medical images, such dataset inconsistencies are often addressed using normalization: A process involving various methods that alter the data to fit within a set of property criteria. Gray-level discretization, an example of a normalization method, clusters pixels based on their intensity, assigning them to one of a set number of gray-level bins [35]. This ensures that all images in a dataset have an equivalent number of intensity levels. After a degree of normalization is performed, many models

appear to achieve increased classification accuracy using the normalized data. Additional benefits from normalization may arise depending on the stage of the process where it is applied: When using a combination of non-texture and gray-level texture features from multicentre datasets, intra-institutional normalization seems to allow for increased performance when compared to normalization of an already combined dataset [36].



(a) (b)

Figure 2.3: The original MRI volume (a) is pre-processed using the FreeSurfer neuroimaging analysis software package with autorecon1 (b). Note the effect of the normalization on structural features: While macro structures are more pronounced (b) there is a decrease in fine texture detail in the grey matter regions when compared to the original. Altogether, the processing involves: 1. Motion correction and conform, 2. NU (Non-Uniform intensity normalization), 3. Talairach transform computation, 4. Intensity Normalization 1

2.3.5 Normalization Concerns

Despite the advantages that normalization methods may bring, they can also introduce additional considerations that decrease the effectiveness of a system relying on texture analysis. The method used for gray-level discretization, for example, has a direct impact on the reproducibility of GLCM texture features [35]. Other normalization techniques may alter image voxels in ways that can cause details present in the original texture to be lost, potentially causing a decrease in the maximum potential of a classification system. In the case of contrast or intensity normalization, image

structure may be enhanced in certain areas to achieve a greater degree of comparability between images. However, this comes inevitably at the cost of structural detail in other areas (Figure 2.3). Image smoothing can also hurt classification performance when using HOG-based methods as a significant portion of image information present in these features derives from abrupt edges at fine scales [12]. Hence, relying on an invariant texture analysis method that is resistant to image differences introduced by scanning discrepancies may achieve superior performance compared to using normalized data with an arbitrary feature extraction method.

2.4 Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis is a neurological disorder with a prevalence of approximately 2.7 cases per 100 000 individuals in Europe and North America [37], and whose patients have survival times varying from over 20 years to a few months [38]. The mean age of offset is between 55 and 66 years of age in the US, Europe and Japan [37, 39], and juvenile cases are seldom observed [38]. However, reversals in disease progression where some motor function is regained are also rare [40]. The illness is also sometimes referred to as Charcot or Lou Gehrig's disease and while usually sporadic, ALS can also be familial-type with a dominant inheritance factor [37]. After onset, a number of symptoms may arise, including involuntary muscle movement, weakness, cramps and function degradation, as well as shortness of breath (dyspnea) and a difficulty swallowing (dysphagia).

2.4.1 The Nervous System

ALS impacts a number of tissues in the brain, which contains both gray and white matter. The brain is a core organ in the central nervous system (CNS) while the overall nervous system is composed of both the CNS and peripheral nervous system (PNS). Regions of the brain are responsible for a variety of functions, for instance, motor components allow for muscle (skeletal, cardiac, smooth) and glandular secretion

control [41]. ALS appears to affect both motor regions, including the motor cortex, and non-motor areas, such as the frontotemporal regions [42]. Degeneration of the white matter, the tissue responsible for signal communication, is most pronounced in the corticospinal tract (CST) and frontal lobes [43].

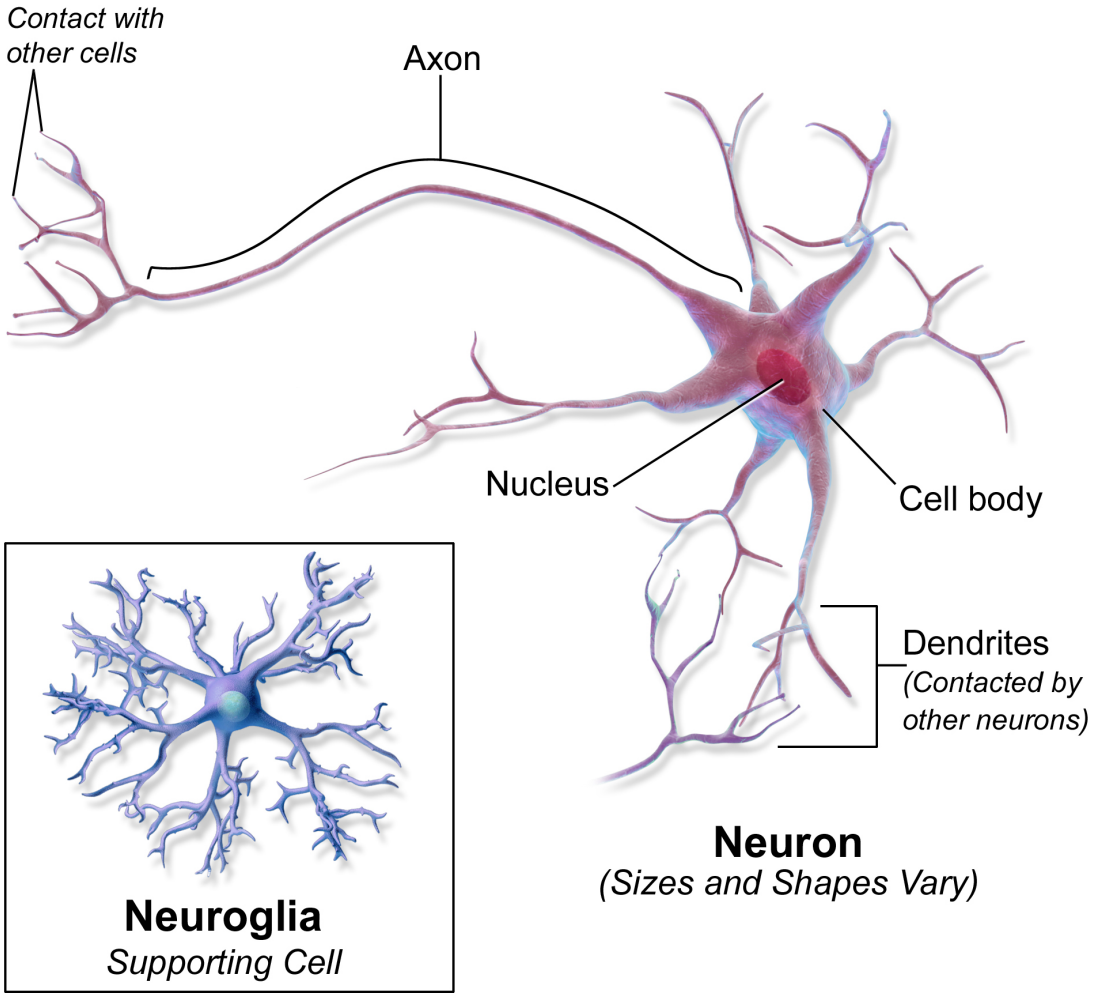


Figure 2.4: Anatomy of a nerve cell (neuron) [44].

2.4.2 Neurons

A basic component of the nervous tissue in these brain sections is nerve cells, also referred to as neurons (Figure 2.4). Neurons allow for signals to travel both to and

from various brain centres, specifically, afferent, or sensory, nerves carry signals to brain while efferent, or motor, neurons innervate muscles and glands or secretory cells. Motor neurons can be further divided into somatic and autonomic, with the first being responsible for movements and the latter for the maintenance of respiration and blood pressure. Another classification for certain motor neurons present in both somatic and autonomic systems, is based on their location. These are known as upper and lower motor neurons, or UMN and LMN, respectively [45]. Upper motor neurons, whose cell bodies are located in the pre-motor and primary motor region of the cerebral cortex, connect to LMNs located in the CNS. If subjected to damage, or lesions, UMNs may cause uncontrolled movement and lower the sensitivity to reflex stimulation. Additionally, increased muscle tightness due to prolonged muscle contraction, or spasticity, may arise. Lower motor neurons, whose cell bodies are located in the brainstem as well as the spinal cord's ventral horn, have their axons extend out of the CNS. Grouped into branchial, visceral, and somatic types, LMNs receive incoming signals from upper motor neurons, interneurons or sensory neurons. Due to LMNs' mediary function in various pathways, lesion symptoms usually involve some form of paralysis [45].

2.4.3 ALS Pathology

Degradation of UMNs and LMNs as well as the degeneration of frontotemporal regions are typical in ALS. While a minority of ALS cases are familial and can be caused by mutations in genes such as C9orf72 [46] or the gene which encodes the antioxidant superoxide dismutase 1 (SOD1) [37, 38], more than 90% of ALS cases are sporadic [47]. Symptoms emerge once axon connections fail and while some neurons are more resistant to the degenerative effects and can even compensate, cell bodies will eventually begin to die once the disease has sufficiently progressed [48]. Clinically, ALS progression is often measured using the ALS Functional Rating Scale, ALSFRS-Revised [49]. This scale is based on a questionnaire used to measure a patient's ability to

perform conventional tasks such as speaking, walking, breathing and salivation with the weighting differing between breathing and activities involving motor functions. Hence, ALSFRS-R has even been adapted for survival modelling [50]. However, the use of the scale relies on an established diagnosis [51], which can be challenging to establish in the first place [52], and may result in the delay of therapeutic intervention.

2.4.4 ALS Treatment

Although the medications Riluzole and Edaravone have been approved by the U.S. Food and Drug Administration for the treatment of ALS, the medications have side effects and generally only slow symptom progression [53]. Nevertheless, there does appear to be an increase in the survival time of some patients, with 49-86% of patients in the US, Europe and Japan receiving Riluzole [39]. Thus, improvements to the duration and accuracy of the diagnostic process, possibly through the introduction of supporting classification systems, may have a significant impact on the survival of ALS patients.

2.4.5 ALS Biomarkers

Currently, prompt diagnosis of patients is hampered by a lack of clear and reliable biological markers (biomarkers) [52], or measures of cell and tissue properties, shifting a substantial portion of the process onto the results of physical examinations [39]. Partially, this can be explained by the uncertainty surrounding how ALS spreads throughout the CNS, with both a focal instantiation point [52, 54] and a multiple hit hypothesis [54] being proposed. Disease progression rate, illustrated through techniques like muscle MRI [55], shows some promise as a biomarker [56] but cannot be used for initial diagnosis. Clinical brain MRI, including T1-weighted, T2-weighted and fluid-attenuated inversion recovery (FLAIR), of ALS patients appears largely normal and there is no biomarker which can be used to evaluate neurodegeneration. Hence, due to the absence of diagnostic evaluations capable of establishing the pres-

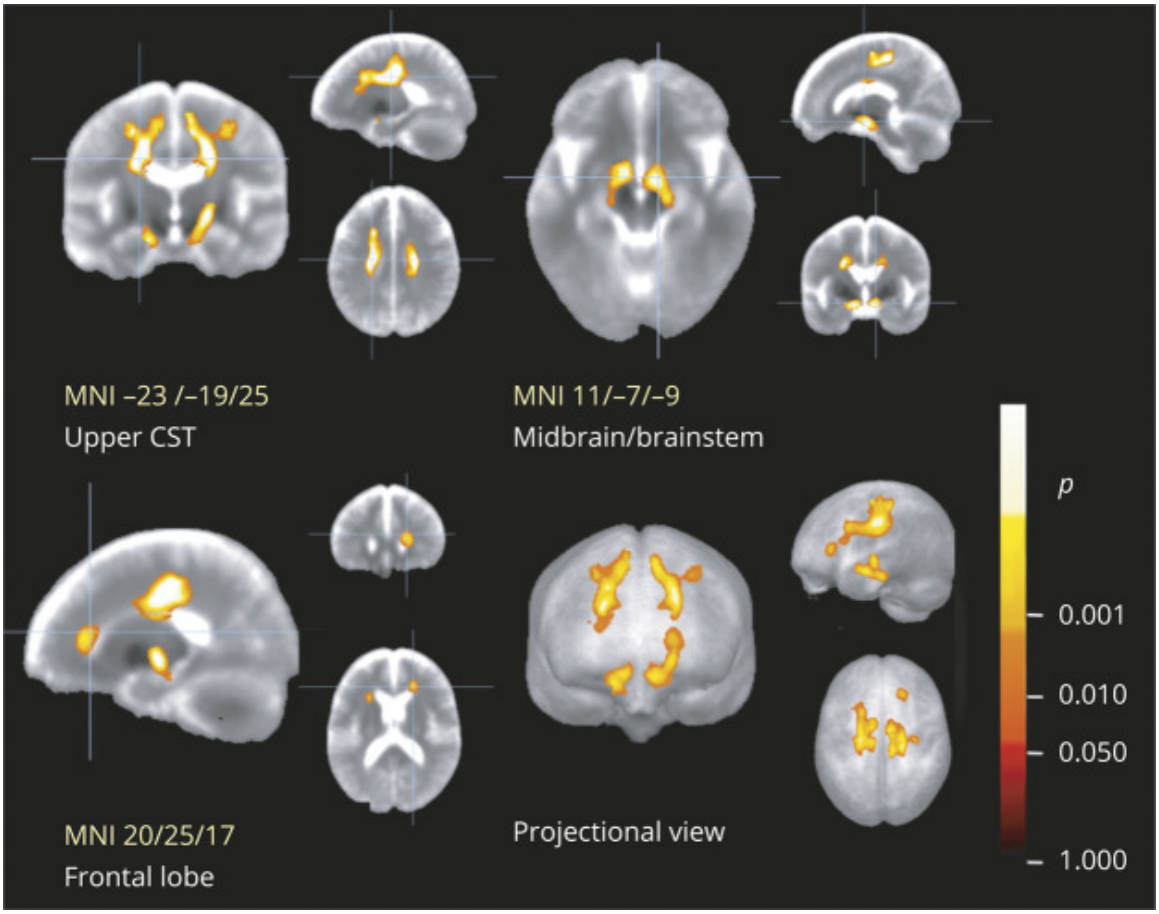


Figure 2.5: ALS patient vs control whole-brain spatial statistics for cross-sectional comparison of fractional anisotropy maps [43]. White matter degeneration appears to occur mostly in the corticospinal tract and frontal lobes.

ence of ALS, various tests are used to rule out other diseases instead. Atrophy of the motor cortex [57] and signal changes in the CST [58] have been observed in T1 and T2/FLAIR images, respectively. However, these observations are of limited sensitivity and specificity, preventing them from being applied clinically. Therefore, texture analysis of these images is required to extract signals of degeneration which are not visible otherwise. Degenerative changes in upper motor neuron function [59] and the CST [60] have already been linked with textures, while further insights derived from a form of brain MRI termed diffusion tensor imaging (DTI) [61] may also be applicable in earlier stages of ALS (Figure 2.5). Possible markers derived from DTI images include detectable changes to motor capabilities in frontotemporal regions of the brain

[62] and more pronounced changes in white matter that composes motor tracts [61] such as the CST [43].

2.4.6 ALS MRI Datasets

One impediment to the investigation of biomarkers in ALS is the relative lack of imaging datasets spanning multiple centres with a degree of coherence in protocol [63]. This causes various studies to use small, single-centre datasets which are prone to biases, resulting in findings that can lack generalizability. With the intention of addressing these concerns, a clinical research platform spanning multiple centres called the Canadian ALS Neuroimaging Consortium (CALSNIC) was created in 2014 [64]. By using a multi-faceted approach including both T1 and T2-weighted MRI sequences, the study aimed to create a more robust and generally representative dataset not subject to as many biases as single centre studies. The first iteration, CALSNIC-1, included 87 ALS patients and 65 healthy subjects from six sites in Canada. CALSNIC-2, the second iteration of the study, now uses an increased sample size and introduces additional sites from Canada and the United States.

2.5 Texture Analysis for ALS Data

Datasamples from the CALSNIC dataset have been applied to evaluate the reliability of 3D texture analysis for MR images of ALS patients. 3D ROI-based extensions of GLCM features, such as VGLCM-3D [65], appear to have the potential to be highly reliable across multiple centres [66]. However, the reliability seems to be strongly dependent on multiple factors, particularly the feature type, region analyzed and whether a method performed an ROI or voxelwise analysis. Specifically, analysis using regions of interest appears to have increased reliability when compared to considering all voxels in a volume, with brain regions included in an ROI impacting intersite reliability. However, ROI based analysis often relies on manual selection or expert input, significantly increasing the involvement of such a study and makes

these methods unattractive for clinical deployment. Additionally, differences to the dataset may be introduced due to contouring variability in manual segmentations. Hence, methods such as VGLCM-TOP-3D, which extend GLCM to define features for 3D voxels instead of 2D pixels, are attractive for usage but can be outperformed by methods that rely on specific ROIs, such as Modified Co-occurrence Histograms of Oriented Gradients (M-CoHOG) [8].

2.5.1 Modified CoHOG

M-CoHOG is a modification of the traditional Co-occurence Histograms of Oriented Gradients feature extraction method that makes a number of changes and adds a feature normalization technique prior to the classification step [8]. A coronal slice from a subject in the CALSNIC-1 dataset is selected by an expert and segmented to create an ROI mask (pre-processing). Additional coronal slices adjacent to the selected slice, specifically 5 in either direction, are also used in conjunction with the segmentation mask.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \tag{2.3}$$

The selected ROI is then scaled, by a factor of 0.33, for example, after which each pixel's gradient orientation is calculated using the Sobel kernel (Equation 2.3) to determine the horizontal and vertical gradient operators. The gradient orientation (GO) is then determined by

$$GO = \arctan\frac{G_y}{G_x}, \tag{2.4}$$

after which the GO pixels can be quantized into the closest bin. In the case of 12 bins, each bin covers a span of 30 degrees in order to cover the full 360 degrees for all directions. Extraction of the texture features is performed using the quantized gradient orientation image as data. Instead of computing co-occurrence matrices (CMs) for image subregions, M-CoHOG calculates a single matrix per offset resulting

in a number of histograms equivalent to the number of offsets, which, in turn, is determined by the radius of the neighbourhood . Therefore, the CM is calculated for each offset x,y, of each pixel p,q, such that $T = [GO(p,q) = i$ and $GO(p+x, q+y) = j$ and $GM(p,q) \geq \tau$ and $GM(p+x, q+y) \geq \tau]$ and

$$CM_{x,y}(i,j) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \begin{cases} 1 & \text{if } T \text{ is True} \\ 0 & \text{Otherwise} \end{cases}, \tag{2.5}$$

where for pixel (p,q), $GO(p,q)$ is the gradient orientation, $GM(p,q)$ is the gradient magnitude and τ is a threshold for the co-occurrence count used to reduce the effect of noise. Next, the matrix is normalized: first, a co-occurrence matrix CMV is calculated using Equation 2.5 but with $\tau = 0$. The normalized matrix CM' is then determined by

$$CM'_{x,y}(i,j) = \frac{CM_{x,y}(i,j)}{\sum_{g=1}^{O} \sum_{h=1}^{O} CMV_{x,y}(g,h)}, \tag{2.6}$$

Finally, the CM' results are concatenated to form the overall 1D feature vector for the whole slice. An overview is provided in Figure 2.6.

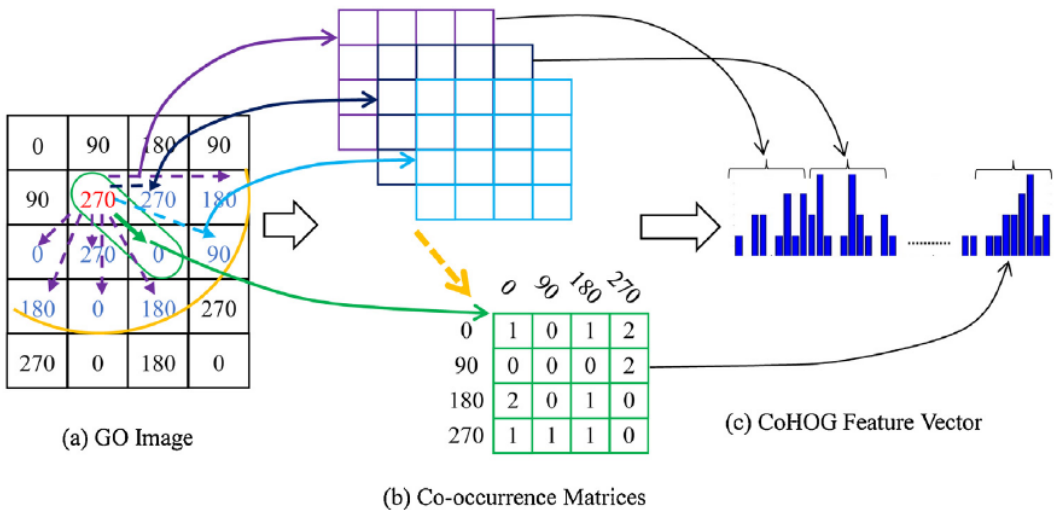


Figure 2.6: Overview of the CM calculation using the GO image as well as the concatenation into the feature vector [8].

2.5.2 Modified CoHOG Classifier

For the classification step, M-CoHOG uses an ensemble classifier that accepts the results of 5 base classifiers as input for an intermediate layer of fully convolutional neural networks (FCNNs), with a final predictor layer receiving the results of all the intermediate classifiers. Two support vector machines (SVMs), with a linear and radial basis function (RBF) kernel, respectively, a K-nearest neighbour (KNN) classifier, a convolutional neural network (CNN), and a random forest classifier compose the base classifier layer. When applied to a multicentre ALS dataset derived from CALSNIC-1, M-CoHOG outperforms both GLCM based and neural network approaches such as VGG16 and ResNet50, suggesting that an ROI-based, CoHOG method is a relatively effective approach for the classification of ALS sMRI images.

Because M-CoHOG requires ROIs that are manually selected and segmented by an expert, the pre-processing requirements of the method are non-trivial. Additionally, some spatial information cannot be captured by the extracted features due to the slice-wise selection of ROIs. Hence, in this thesis we present a volumetric feature extraction method based on M-CoHOG which does not require an expert to select ROIs and is able to function on volume imaging data.

Chapter 3

Methodology

3.1 Accelerating M-CoHOG with CUDA

The original M-CoHOG implementation performs feature extraction using a single CPU thread, making it inefficient for imaging data. However, because the operations involved in the CoHOG method are not dependent on each other's intermediate results, the process can be performed in parallel. We use three CUDA kernels to perform the steps with the highest performance cost on a GPU with support for the CUDA language. Specifically, we use CUDA to accelerate the computation of the gradient orientation image with a Sobel kernel, the calculation of valid pairs for every offset, as well as the calculation of the co-occurrence matrices (Figure 3.1). Calling the compiled kernels is enabled by the mexcuda [67] extension of the Matlab mex function, which is designed to allow for the calling of C/C++ programs or Fortran subroutines from within Matlab.

3.1.1 Gradient Parallel Implementation

To allow for parallel execution of the Sobel operations and gradient calculation, first, the number of threads per block, T, is set to be the image height - 2. The - 2 offset is necessary as the Sobel kernel cannot be applied to boundary pixels. The blocks per grid, B, are then determined by $B = \frac{N+T}{T}$, where N is the total number of elements in the gradient orientation image. The Sobel function is then called for B blocks,

each composed of T threads. Hence, within each thread, the current pixel in the 1D data grid is determined using the x dimension of the block in conjunction with the indices of the current block and thread.

3.1.2 Co-occurrence Matrix Parallel Implementation

For the valid pair count and CM calculation functions, the steps necessary for parallel execution are equivalent to those described for the Sobel and gradient orientation calculation with the exception that T is set to be the full gradient orientation height. A notable point about the implementation of the pair count and CM functions is that they require the use of an atomic addition operation when a value must be added to the result arrays. The reason for this is that a race condition can arise where multiple threads attempt to write to the same location in memory. Because the operations are simultaneous, the result of each thread is likely to be out of order, causing the last value written not to be equal to the total sum expected.

3.2 Extending M-CoHOG to support Volume Data

The volumetric texture analysis method we propose, referred to as Volumetric Co-occurence Histograms of Oriented Gradients (V-CoHOG), is based on CUDA-accelerated M-CoHOG and still shares many similarities with the original method; most additions and alterations directly relate to facilitating the processing of volume information, as opposed to individual slices. We leverage runtime improvements provided by GPU acceleration to allow volume data with substantially more voxels than individual slices to be processed with large neighbourhood sizes. However, CUDA acceleration is only implemented for the valid pair count, co-occurrence matrix and feature vector generation steps as these sections see the greatest performance gain because of parallelization in the M-CoHOG implementation. Specifically, blocks per grid, B, is altered to be a 2D vector consisting of the GO axial size and GO coronal size to improve the facilitation of 3D data while threads per block, T, remains as being equivalent to the

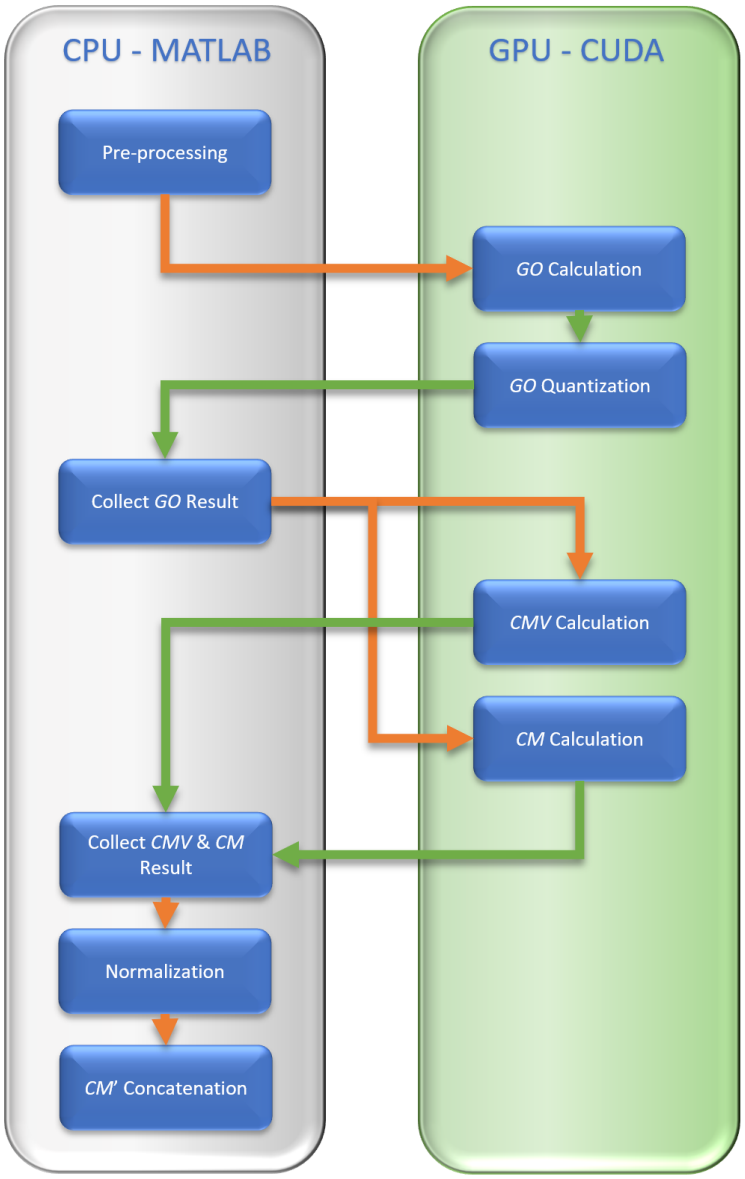


Figure 3.1: Overview of the CUDA-accelerated implementation of M-CoHOG. GO and CM describe the gradient orientation and co-occurrence matrix, respectively. The M-CoHOG method is described in Section 2.5.1. Initial pre-processing of the input data is not GPU-accelerated but once complete, the processed ROI is used to calculate the gradient orientation as well as the quantization of gradient angles in bins. After the result is returned, another CUDA kernel calculates the total number of valid pairs for an offset. Next, the co-occurrence matrix is calculated using another CUDA kernel. Normalization of the co-occurrence matrix occurs in Matlab. Finally, the co-occurrence matrices for all offsets are concatenated into the feature vector. Details pertaining to the CUDA adaptation for the gradient orientation calculation and quantization are described in Section 3.1.1. The valid pair count, co-occurrence matrix calculation and feature vector generation CUDA details are explained in Section 3.1.2.

GO sagittal size.

3.2.1 General Preprocessing

Rather than rely exclusively on manual segmentation masks, we use Brain Extraction Tool (BET) [68] to generate skull-stripped versions of scans in the dataset in conjunction with the corresponding segmentation masks. For the preprocessing step, the original sMRI volume and the BET segmentation mask are scaled by a factor such as 0.33 before the volume is cropped to the size of the minimum bounding box of the mask volume. Hence, once preprocessing is complete, the image and mask volumes are equivalent in size.

3.2.2 Gradient Orientation Calculation

Similar to M-CoHOG, gradient orientations of the input data are calculated using a Sobel operation. Specifically, for every voxel p in image volume V where $p = V(i, j, k)$, Axial (G_A), Sagittal (G_S) and Coronal (G_C) gradients are calculated using 3D Sobel kernels S_a, S_s, S_c, respectively (example Figure 3.5). For example, for G_A:

$$S_a[-1] = \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix} S_a[0] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} S_a[+1] = \begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix} \tag{3.1}$$

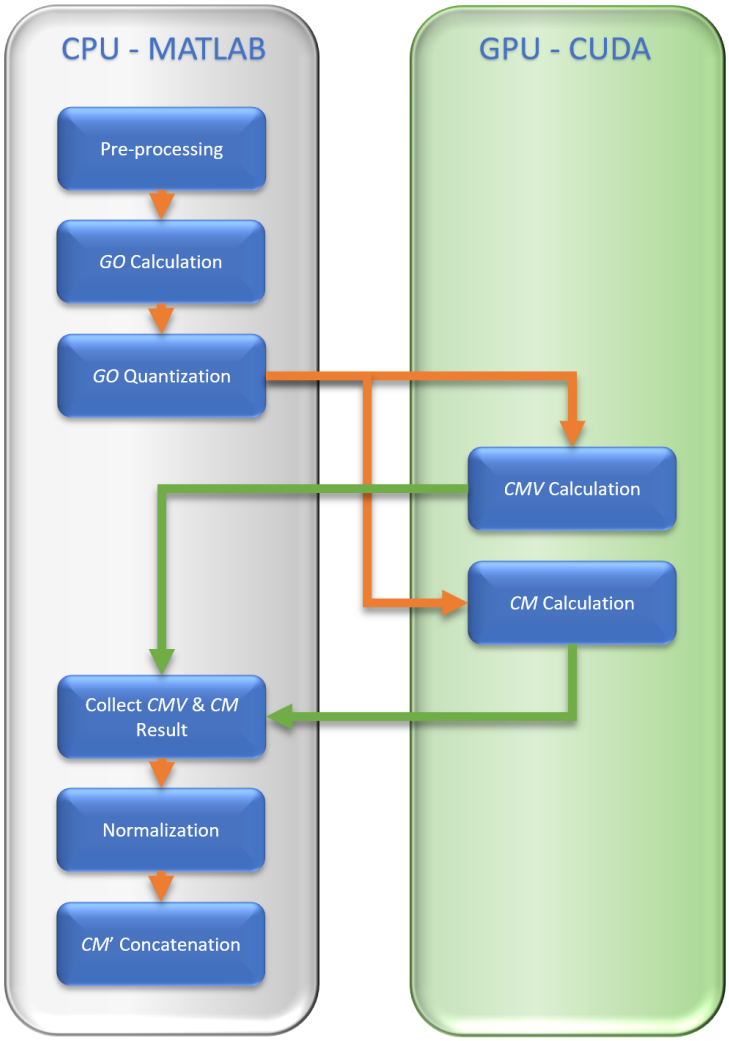
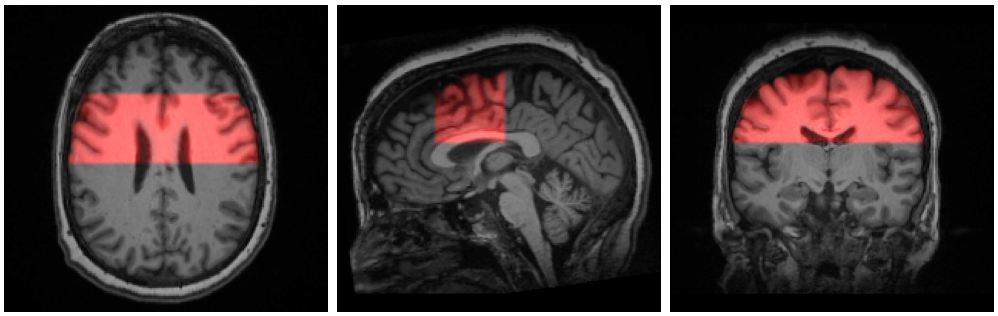


Figure 3.2: Overview of the CUDA-accelerated implementation of V-CoHOG. *GO* and CM describe the gradient orientation and co-occurrence matrix, respectively. Pre-processing is described in Section 3.2.1. The calculation and quantization of the gradient orientation is described in Section 3.2.2. Calculation of the co-occurrence matrices is described in Section 3.2.3. Finally, the normalization and concatenation steps are explained in Section 3.2.4.



(a) Axial (b) Sagittal (c) Coronal

Figure 3.3: Region of interest extracted from the CALSNIC1 Centre 1 dataset for the single-centre experiments.

with which we can determine L, M, N using the Hadamard product $\circ$:

$$L = S_a[-1] \circ \begin{bmatrix} V(i-1,j-1,k-1) & V(i-1,j,k-1) & V(i-1,j+1,k-1) \\ V(i-1,j-1,k) & V(i-1,j,k) & V(i-1,j+1,k) \\ V(i-1,j-1,k+1) & V(i-1,j,k+1) & V(i-1,j+1,k+1) \end{bmatrix}$$

$$M = S_a[0] \circ \begin{bmatrix} V(i,j-1,k-1) & V(i,j,k-1) & V(i,j+1,k-1) \\ V(i,j-1,k) & V(i,j,k) & V(i,j+1,k) \\ V(i,j-1,k+1) & V(i,j,k+1) & V(i,j+1,k+1) \end{bmatrix}$$

$$N = S_a[+1] \circ \begin{bmatrix} V(i+1,j-1,k-1) & V(i+1,j,k-1) & V(i+1,j+1,k-1) \\ V(i+1,j-1,k) & V(i+1,j,k) & V(i+1,j+1,k) \\ V(i+1,j-1,k+1) & V(i+1,j,k+1) & V(i+1,j+1,k+1) \end{bmatrix} \tag{3.2}$$

such that, G_A of pixel p at $V(i,j,k)$ is determined by

$$G_A = \frac{\sum_{i=1}^{3} L(i,j) + M(i,j) + N(i,j)}{b} \tag{3.3}$$

where b is the number of angle bins. The Hadamard product of two matrices A and B with equivalent dimensions can be described as: $(A \circ B)_{ij} = (A)_{ij}(B)_{ij}$. The gradient value G of p is then determined as

$$G = \sqrt{G_A^2 + G_S^2 + G_C^2} \tag{3.4}$$

After which G is compared to a threshold τ where

$$G = \begin{cases} G, & \text{if } G \geq \tau \\ 0, & \text{if } G < \tau \end{cases} \tag{3.5}$$

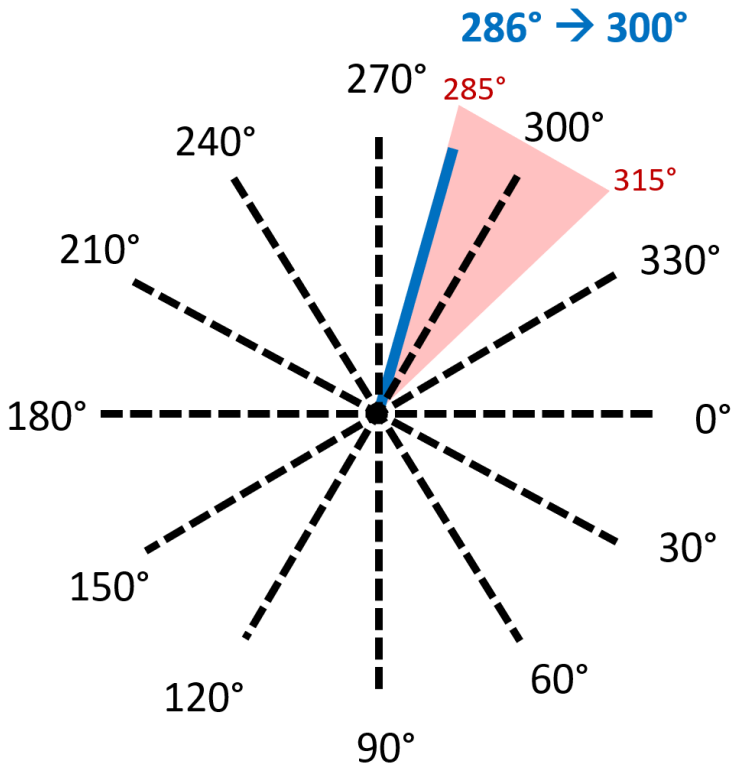


Figure 3.4: V-CoHOG: Illustration of the quantization of the angle 286° into the bin of 300° when given a number of bins parameter equal to 12.

and the optimal value for the threshold τ is determined using a grid search. Hence, if $G > 0$, the gradient orientation GO_{x-y} with $x - y$ corresponding to Axial-Sagittal (A-S), Axial-Coronal (A-C) and Sagittal-Coronal (S-C), is calculated:

$$GO_{x-y} = \text{CalculateOrientation}(\frac{180 \times \arctan(G_x, G_y)}{\pi}, b) \tag{3.6}$$

In CalculateOrientation, the orientation of GO_{x-y} is determined by quantizing the first function parameter in Equation 3.6 into the nearest of b bins (example Figure 3.4).

3.2.3 Co-occurrence Matrix Calculation

Prior to the CM calculation, the total valid pairs (CMV), where pairs are for every voxel and its corresponding offsets within a radius, are calculated per offset using a threshold $\tau = 0$. The co-occurrence matrix CM is determined as per Algorithm 1

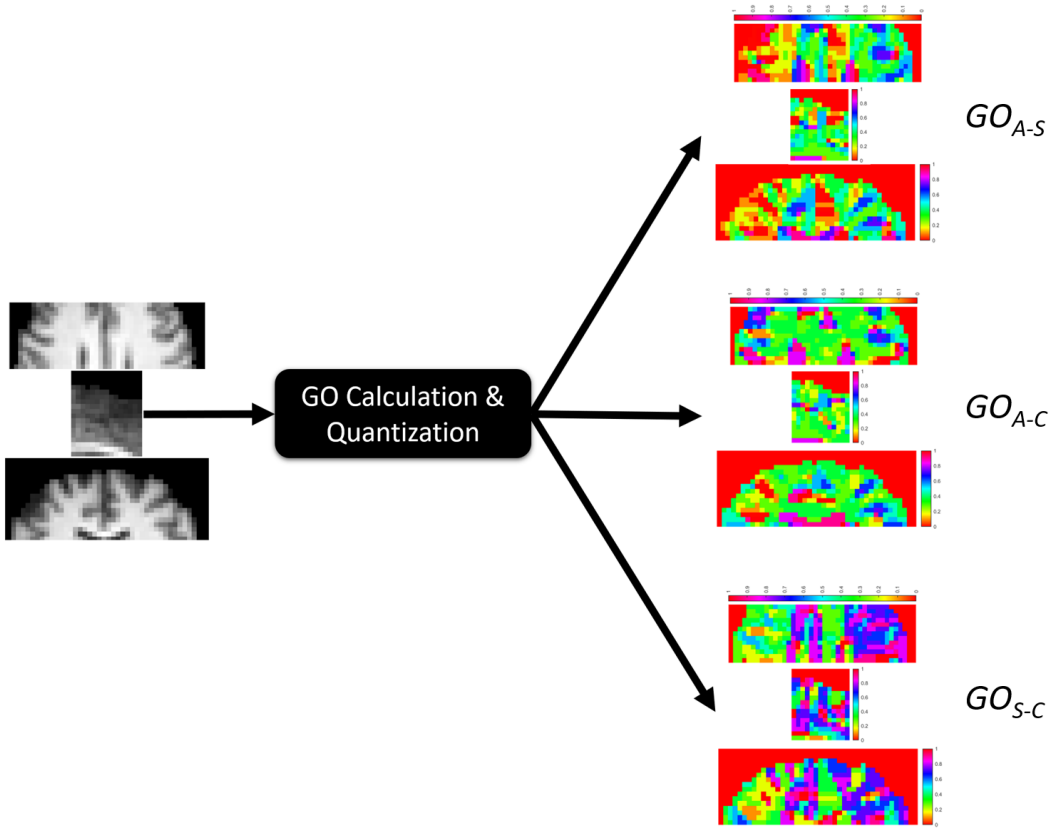


Figure 3.5: V-CoHOG: Graphical representation of the result of the gradient orientation calculation and quantization steps. The resulting orientations are displayed using a color map corresponding to the range 0 - 360°

where the CM calculation follows the plane p+x,q,s+y. The other two planes are shown in Figure 3.6. In Algorithm 1, H is used to denote a matrix containing all CMs, specifically, for a given offset x and all i and j, $H(i, j, x)$ corresponds to the CM for offset x.

Algorithm 1 V-CoHOG: calculation of CMs.

for $p \leftarrow 1$ to sizeof(GO_{Axial}) **do**
  for $q \leftarrow 1$ to sizeof($GO_{Sagittal}$) **do**
    for $s \leftarrow 1$ to sizeof($GO_{Coronal}$) **do**
      if $GO(p, q, s) \geq \tau$ **then**
        $i \leftarrow 1$+idivide($GO(p, q, s)$,round($\frac{360}{b}$))
        $OffsetNumber \leftarrow 1$
        for $x \leftarrow OffsetRadius$ to 0 by -1 **do**
          if $x = 0$ **then**
            $r \leftarrow 1$
            $e \leftarrow OffsetRadius$
          else
            $r \leftarrow OffsetMask(x, 1)$
            $e \leftarrow OffsetMask(x, 2)$
          end if
          for $y \leftarrow r$ to e **do**
            if $(p + x \leq$ sizeof($GO_{Sagittal}$)$\&p + x > 0)\&(s + y \leq$ sizeof($GO_{Coronal}$)$\&s + y > 0)$ **then**
              ▷ Check whether the second voxel is inside the image
              if $GO(p + x, q, s + y) \geq \tau$ **then**
                $j \leftarrow 1$+idivide($GO(p + x, q, s + y), round(\frac{360}{b})$)
                $H(i, j, OffsetNumber) \leftarrow H(i, j, OffsetNumber) + 1$
              end if
            end if
            $OffsetNumber \leftarrow OffsetNumber + 1$
          end for
        end for
      end if
    end for
  end for
end for

3.2.4 Co-occurrence Matrix Normalization

Normalization of the Co-occurrence matrix using the total number of valid pairs for each offset (x, y) is performed in accordance with the M-CoHOG method [8] in

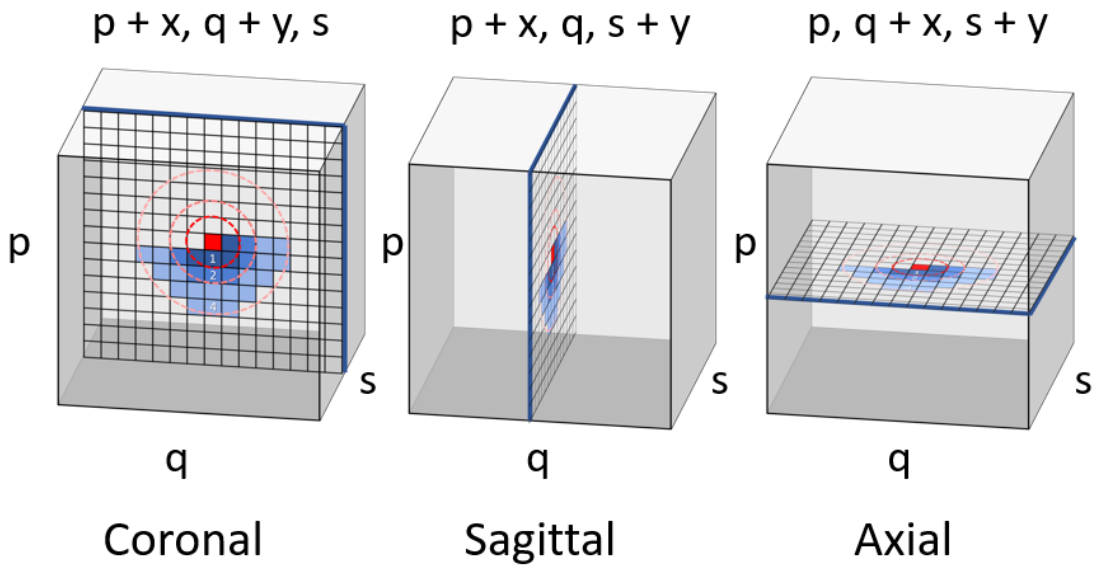


Figure 3.6: V-CoHOG: The three planes along which the co-occurrence matrix can be calculated.

order to increase the stability of features in light of noise and illumination changes. Hence, the normalized co-occurrence matrix CM' is determined by Equation 2.6 where CMV is a co-occurrence matrix calculated using Algorithm 1 with a threshold $\tau = 0$. Furthermore, CMV has a dimensionality that is one less than CM as the pairs are summated per offset without consideration for any orientation bins. After all CMs are normalized they are concatenated into the final feature vector (Figure 3.7) whose size is determined by

$$\text{sizeof(Feature Vector)} = N \times \text{sizeof}(CM') \tag{3.7}$$

where N is the number of offsets.

3.3 Feature Selection

While features derived from texture analysis have proven to be an effective tool for various computer vision tasks and GPUs are practical for the acceleration of their derivation, the quantity of features computed from a single image may be overwhelm-

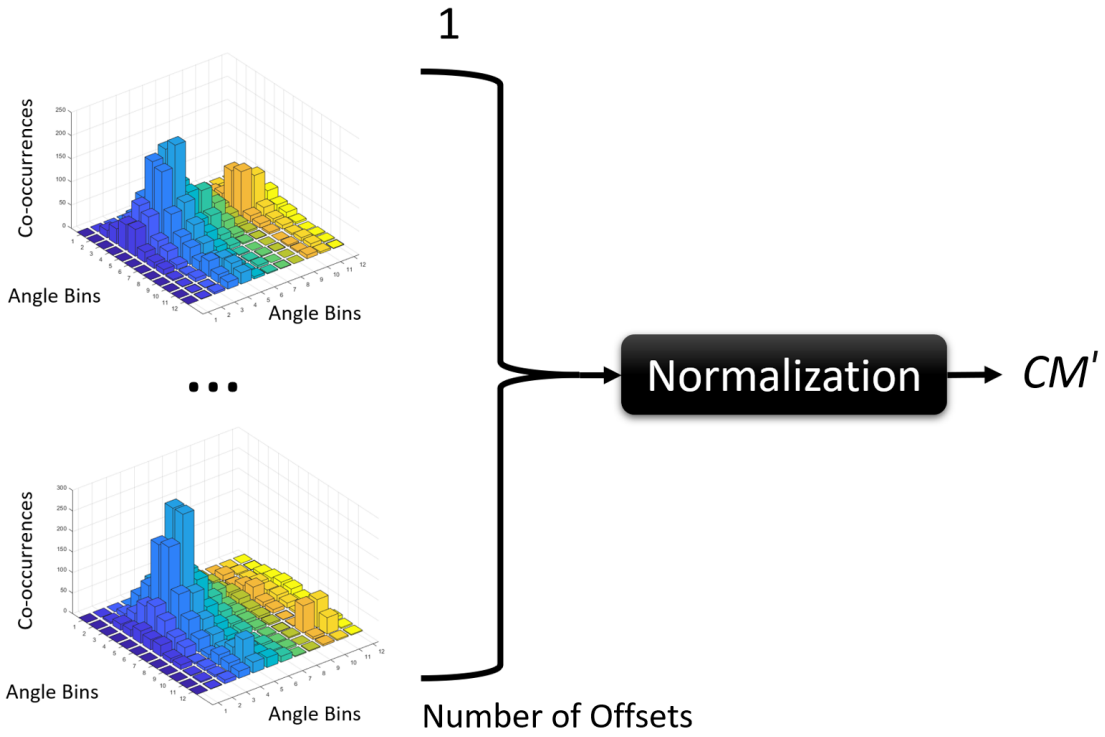


Figure 3.7: V-CoHOG: Overview of the normalization and concatenation of the CMs into the final feature vector.

ing for certain applications. The number of features used to represent a single dataset, or feature dimensionality, is usually reduced through two common methods: Feature projection and feature selection. Projection is often performed using Principal Component Analysis as well as related methods. Feature selection methods can be grouped into a handful of general categories including filter, wrapper, embedded, and structured methods [69]. Filter methods base their selection criteria on a performance measure and include methods such as information gain, chi-square, minimum redundancy maximum relevance (mRmR) [70] and Relief [71]. On the other hand, wrapper methods select features based on their performance on a modelling algorithm used as a black box evaluator and embedded methods select features during execution of the modelling algorithm. Finally, structured methods involve specific algorithms tailored towards selecting features with a considerable amount of independence.

3.3.1 Relief

Relief [71] is a filter method for feature selection that was developed to address some of the shortfalls of inductive, or observational, learning. Specifically, older methods relying on metrics such as information gain assume that features are conditionally independent; hence, the features are expected to be in domains with strong conditional dependencies between each other [21]. On the other hand, Relief is not as sensitive to being misled by feature interaction and is more noise-tolerant [71]. This is due to how Relief estimates attributes according to how well their values distinguish between instances that are in proximity to each other.

3.3.2 ReliefF

Relief estimates feature weights based on their ability to distinguish between dataset instances that are in proximity to each other [63]. The Relief filter method has since been extended via ReliefF [72] to allow the method to be applied to incomplete datasets as well as problems with more than two classes. ReliefF estimates probabil-

ities with greater reliability than Relief using a parameter K to refer to the number of neighbours involved with the scoring update process [73].

Because increases in the radius size or number of bins used by V-CoHOG cause an increase in the total number of features, the feature set used for classification needs to be optimized to remove features which have no discriminate characteristics for the classification task. Hence we utilize ReliefF with Matlab to generate a feature ranking of the V-CoHOG feature vector. Specifically, we use a random subset of the training and testing data equal to 80% of the total dataset size and split the subset into five folds. ReliefF is then used to generate five feature rankings for the subset with a different fold excluded for each ranking computation. Finally, the top N features from every fold are selected after which any features not ubiquitous to all five feature ranking subsets are removed. Furthermore, in some of the ALS experiments, the selected features are also used to generate segmentation maps of the gradient orientation volume to indicate which areas of the ROI are used for the classification task.

3.4 Classifier

A stacking ensemble classifier with 3 levels based on M-CoHOG [8] is used for classification of the feature vectors generated using the V-CoHOG methodology (Figure 3.8). The results of each layer are used as input for the next, with the intermediate and predictor layers being composed solely of FCNN nodes.

3.4.1 Support Vector Machines

Support vector machines (SVMs) are a type of learning machine, or classifier, originally intended for two-group, or binary, classification problems [74]. As input, the classifier takes input vectors of sample features and maps them non-linearly to a high dimensional space via a kernel function. Within the same high dimensional space, a decision surface that separates the data into two classes is determined. This optimal

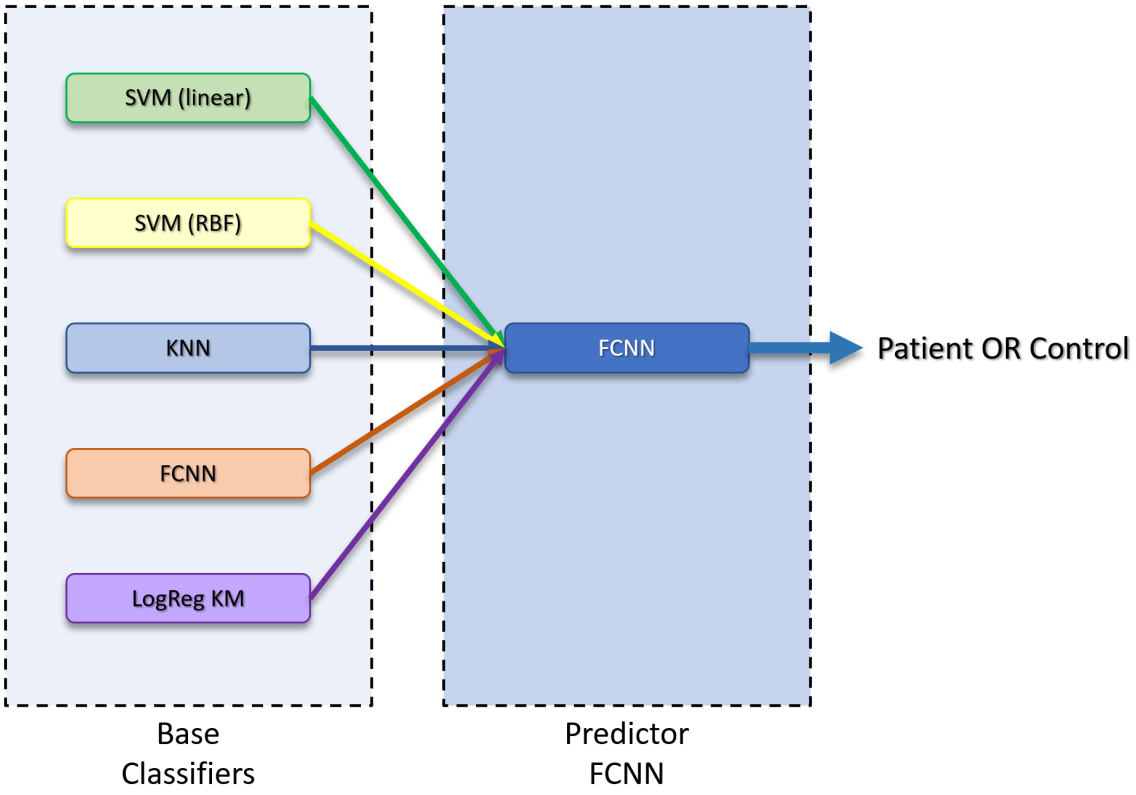


Figure 3.8: Ensemble stacking model used for the classification of the feature vectors.

hyperplane discriminates the training data with a maximal margin. The decision rule used by SVMs, which involves solving a quadratic optimization problem, also relies on the elements of the training set as well as a function which determines the convolution of dot-products. This function is referred to as a kernel, K(u, v), and can be altered to allow for non-linear decision surfaces, such as with a radial basis function (RBF). Because SVMs yield good classification performance when compared to other convex optimization methods [75], they are a popular method outside of ensemble models.

3.4.2 K-Nearest Neighbour

K-nearest neighbour (KNN) is a non-parametric classification technique, which involves assigning class membership to a data instance based on a vote of its K nearest neighbours [76]. Hence, K is a parameter for whom the best value depends on the dataset. To determine the nearest neighbours, distance metrics such as Mahalanobis,

Minkowski distance exponent or Euclidean distance can be used.

3.4.3 Fully Convolutional Neural Network

Fully convolutional neural networks (FCNNs) are a classifier composed of layers that attempts to mimic the approximate function of visual cortices in animals [77]. In particular, the network has an input layer, one or more hidden layers, and an output layer and is a type of convolutional neural network (CNN) in which all layers are learning filters and there are no fully connected layers; i.e. layers where every neuron in one layer is connected to a neuron in another layer. The convolution operations are performed in the hidden layers and generally involve the calculation of the dot product of the incoming matrix and a convolution kernel.

3.4.4 Logistic Regression Kernel Machine

While SVMs are the most well-known example of kernel machines, a variety of different learning methods may be combined with the dimensionality expansion implemented by kernel functions [78]. Logistic regression is a statistical method to model an equation by fitting the output values between 0 and 1 via the logistic function [79]:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.8}$$

Hence, after mapping the features of a dataset to a higher dimensional feature space, a logistic regression classifier can be trained and then applied to a testing dataset for a binary classification task.

3.4.5 Ensemble

Classification methods relying on the output of multiple systems often outperform individual classifiers [80]. This is because individual classification systems may have comparable training performance but divergent generalization performance. Hence, considering the results of multiple classifiers can reduce the risk of relying on a single method that happens to be ill-suited for a given dataset. While there are different

ways in which an ensemble system can arrive at a consensus, a stacking architecture was selected: Stacking ensembles involve a meta-learning algorithm that optimizes the predictions from a set of base classifiers into a single prediction.

Chapter 4

Experimental Results

In this chapter, we discuss the results obtained using the previously described methods for a variety of datasets, with a particular focus on MRI imaging data of ALS patients. While most techniques were implemented in Matlab, GPU acceleration was performed via calls to execute CUDA kernels. Specifically, any runtime-performance measuring tests were performed on a system with a Broadwell Intel Xeon 6 Core CPU, 16GB RAM and two Nvidia Quadro M4000 graphics cards. Nevertheless, CUDA operations were executed on a single GPU.

While we use imaging data from CALSNIC-1 and CALSNIC-2 for both runtime and classification experiments, we also used two other datasets in order to demonstrate that our methods are not strictly applicable to ALS imaging data. We generated a number of feature maps that provide insight into the origin of selected features used for classification in addition to the classification results themselves. As a comparison, we consider the classification results of 4 3D CNN based methods for the MRI imaging data. This also provides us with a reference point for the CALSNIC-2 dataset for which no previous classification results have been published.

4.1 Accelerating M-CoHOG with CUDA

In this section, we discuss the results obtained with our GPU-accelerated implementation of M-CoHOG in terms of a comparison between execution using only the CPU,

as is the case in the original method, and with the use of CUDA kernels for the main sections of the algorithm. The comparison involved a single parameter variation for both datasets in order to demonstrate the potential impact a single parameter may have on the runtime requirements and how the impact differs when the method is executed wholly on the CPU or not.

4.1.1 Datasets

In order to evaluate the runtime improvements yielded by the CUDA-accelerated implementation of M-CoHOG, we used the Edmonton Centre (Centre 1) subset of the CALSNIC-1 dataset, which included 38 MRI volumes, as well as a subset of the DIV2K dataset [81, 82], which included a selection of 400 2K resolution images. For the MRI data, individual coronal slices were extracted from the volume and processed as the M-CoHOG method does not support the generation of feature descriptors for volume data. While the ALS data was homogenous in terms of subject matter, the DIV2K dataset included images of various places and objects (4.1).

4.1.2 Runtime

Table 4.1: CUDA-Accelerated M-CoHOG runtimes using slices sourced from CALSNIC MRI data. Values are reported as averages over 10 runs.

Centre	Scaling	Sobel & GO	CM	Std Dev.	Avg. Runtime (ms)
1	0.33	CPU	CPU	71.9	9227
1	0.33	CUDA	CPU	57.6	8763
1	0.33	CPU	CUDA	48.1	6139
1	0.33	CUDA	CUDA	39.3	5647
1	1.00	CPU	CPU	106.6	39446
1	1.00	CUDA	CPU	52.2	34732
1	1.00	CPU	CUDA	82.7	11420
1	1.00	CUDA	CUDA	58.4	6476

Figure 4.1: A selection of images from the DIV2K dataset. In total, 400 pictures were selected in order to evaluate runtime performance of the M-CoHOG feature descriptor generation method.

Table 4.2: CUDA-Accelerated M-CoHOG runtimes using images from the DIV2K dataset. A scaling factor of 0.25 was applied to all images during the pre-processing step. Values are reported as averages over 10 runs.

Dataset	Radius	Sobel & GO	CM	Std Dev.	Avg. Runtime (ms)
DIV2K	1	CPU	CPU	346.7	165290
DIV2K	1	CUDA	CPU	333.0	140581
DIV2K	1	CPU	CUDA	93.2	40570
DIV2K	1	CUDA	CUDA	42.5	16069
DIV2K	2	CPU	CPU	5470.7	321678
DIV2K	2	CUDA	CPU	930.3	295830
DIV2K	2	CPU	CUDA	80.3	40726
DIV2K	2	CUDA	CUDA	35.1	16150

We observed a decrease in overall runtime in all cases where CUDA kernels were used in order to execute sections of the M-CoHOG algorithm (Table 4.1, 4.2). While the improvements were less extensive for the MRI data (Table 4.1), the DIV2K data could be processed up to 19.9× as quickly when using a neighbourhood radius of 2 (Table 4.2). Unfortunately, the further acceleration of the runtime would likely benefit the most from a reduction in pre-processing time, which cannot be improved as trivially due to the reliance on a multitude of imaging functions. Overall, the performance benefits of the GPU accelerated implementation were less pronounced when using smaller images and neighbourhood sizes where the advantages of GPU thread count cannot be fully utilized.

The performance improvements we observed seem to be similar to those achieved by other attempts to parallelize the generation of HOG features such as [15] who found their GPU method to allow 16.8× faster execution in a real-time moving human detection system. Another method which saw the use of GLCM feature descriptors instead of HOG for abnormality detection in mammograms found a GPU accelerated approach to allow for up to 19.22× faster runtimes depending on the block size used [21]. Hence, GPU acceleration continues to be a valid approach to decreasing the runtime requirements of HOG-derived feature generation methods.

4.2 ALS Patient Classification

In this section, we discuss the results obtained when classifying various subsets of the CALSNIC-1, as well the CALSNIC-2 datasets, using the proposed methodology for V-CoHOG. We use four centres from the CALSNIC-1 dataset in both single and multicentre combinations while the CALSNIC-2 dataset is used without partitioning. The effects of various parameters on the classification results are discussed for the Edmonton (Centre 1) dataset. All V-CoHOG features are calculated using a threshold $\tau = 3$ (Algorithm 1) and features are selected using ReliefF with $K = 6$ (Section 3.3.2). In addition to the classification results, we generated feature maps to help

identify selected features' areas of origin and used 3D CNN methods in order to obtain baseline results for comparison.

4.2.1 Evaluation Metrics

In addition to accuracy (Acc), standard deviation (Std Dev) and variance (Var) metrics, we report the sensitivity (Sens), specificity (Spec) and F-score as averages over 1000 runs. Sensitivity is defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4.1}$$

where TP is the number of true positives and FN is the number of false negatives. Specificity is defined as

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4.2}$$

where TN is the number of true negatives and FP is the number of false positives. The F-Score used is the harmonic mean of the sensitivity and specificity, and is defined as

$$\text{F-Score} = \frac{2 \times \text{Sensitivity} \times \text{Specificity}}{\text{Sensitivity} + \text{Specificity}} \tag{4.3}$$

For every iteration, a random 80-20 train-test split was used and the number of features reported corresponds to the amount used for classification post selection. When discussing the results for particular data samples, the number of correct classifications compared to the total appearances of that particular data sample in the test set were recorded. Results reported for CNN methods used as a comparison were obtained by averaging the metrics achieved across 5 folds in 1 run per fold.

4.2.2 Datasets

The CALSNIC-1 and CALSNIC-2 datasets (Table 4.3) both saw the implementation of four directives aimed at optimizing the comparability of the collected MRI data across sites [64]. These included the exclusive use of 3 T MRI systems; the use of only vendor-supplied sequences on these systems; sequence parameter adjustments with assessments of image properties, such as signal-to-noise ratio, for improved

comparability; and the use of fixed spatial parameters across vendors when viable. Nevertheless, there remain certain inter-vendor differences, as some subjects would not fit into the preferred 64-channel coil and required the use of alternate coil arrays. For CALSNIC-1, centres 1 and 4, and centres 2 and 3, shared the MRI system vendors Siemens and GE, respectively. Only T1-weighted scans were considered for classification as previous applications of texture analysis to T1-weighted images have provided evidence of cerebral degeneration [60, 83].

Subjects labeled as patients underwent evaluation with a more robust set of clinical assessments than is typically recommended, including the use of the revised ALS functional rating scale (ALSFRS-R), forced vital capacity (FVC), UMN and LMN function examination, as well as cognitive and behavioural tests [64]. Generally, the included patients needed to be diagnosed with either sporadic or familial ALS and meet the revised El Escorial research criteria [84].

Table 4.3: Patient and control numbers in the various centres of the CALSNIC-1 and CALSNIC-2 datasets as well as the MRI acquisition properties used for classification experiments. The CALSNIC-2 dataset was not divided into centres. Total, Res and Plane correspond to the total subjects, resolution and acquisition plane, respectively.

Centre	Patients	Controls	Total	Type	Res (mm)	Plane
Edmonton (1)	23	15	38	T1-weighted	1x1x1	Axial
Toronto (2)	22	12	34	T1-weighted	1x1x1	Axial
Calgary (3)	11	8	19	T1-weighted	1x1x1	Axial
Montreal (4)	13	7	20	T1-weighted	1x1x1	Axial
CALSNIC-1	69	42	111	T1-weighted	1x1x1	Axial
CALSNIC-2	113	112	225	T1-weighted	1x1x1	Sagittal

4.3 Single Centre

In this section we discuss the impact of various parameters on the classification performance of V-CoHOG as well as the classification results for the four centres in

Table 4.4: Parameters found to yield the optimal classification results for the ALS datasets after a parameter search. A centroid computation was used to determine the axial offset for the Centre 1-4 combined and CALSNIC-2 datasets. GO types A-S, A-C, S-C correspond to Axial-Sagittal, Axial-Coronal and Sagittal-Coronal, respectively.

Centre	1	2	3	4	1 - 4	CALSNIC-2
Scaling	0.33	0.33	0.33	0.33	0.33	0.33
Coronal Offset	0.52	0.56	0.52	0.53	0.48	0.44
Coronal Slices	15	15	15	15	11	13
Axial Slices	18	18	18	18	Centr+9	Centr+5
Radius	3	3	3	3	3	3
Bins	12	12	12	12	12	12
GO Type	A-C	A-C	A-C	A-C	A-C	S-C
CM Plane	Sagittal	Sagittal	Sagittal	Sagittal	Axial	Axial
Features	258	174	186	164	268	71

Table 4.5: Ensemble classifier results for the ALS datasets. Results from M-CoHOG are included for comparison. Note that while a 80-20 train-test split was used for the proposed method, M-CoHOG used a split of 70-30, respectively. The accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Centre	Method	Acc	Sens	Spec	Std Dev	Var	F-Score
1	M-CoHOG	0.795	0.775	0.814	-	-	0.794
1	V-CoHOG	0.860	0.889	0.812	0.118	0.014	0.848
2	M-CoHOG	0.766	0.726	0.804	-	-	0.763
2	V-CoHOG	0.792	0.874	0.627	0.147	0.022	0.730
3	M-CoHOG	0.732	0.686	0.785	-	-	0.732
3	V-CoHOG	0.791	0.868	0.715	0.209	0.044	0.784
4	M-CoHOG	0.716	0.695	0.737	-	-	0.715
4	V-CoHOG	0.870	0.895	0.792	0.167	0.028	0.841
1 - 4	M-CoHOG	0.745	-	-	-	-	-
1 - 4	V-CoHOG	0.770	0.820	0.684	0.075	0.006	0.746

the CALSNIC-1 dataset. Generally, the single centre experiments have dataset sizes which are in line with those used by many other studies that propose ALS classification methods, often not exceeding a few dozen samples.

4.3.1 Centre 1

Centre 1 contained the most samples from the CALSNIC-1 dataset and was, therefore, selected for the initial parameter search. As shown in Table 4.4, 9 of 10 parameters effectuated the generation of the feature descriptors while the number of selected features was determined later during the feature selection step. In all the parameter search experiments, only the parameter in question was altered while the remainder was set to the value reported in Table 4.4. The cases where no feature selection occurred are an exception to this with the corresponding number of features described in the respective section.

For most of the parameters shared by the original M-CoHOG method as well as V-CoHOG, the values resulting in the greatest classification performance were the same. Specifically, this was the case for the scaling factor and number of quantization bins. This finding increased our expectancy that V-CoHOG's classification performance on the CALSNIC-1 dataset, including individual centres, should be comparable to the state-of-the-art achieved by M-CoHOG [8]. While M-CoHOG used a 70-30 train-test split as opposed to our 80-20, the latter split was selected to increase the comparability of our method with the 5-fold evaluation used for the 3D CNN approaches that also do not involve ROI selection by an expert. We found that our method outperforms M-CoHOG for Centre 1 with a classification accuracy of 86.0% (Table 4.5) as opposed to the latter's 79.5%.

ROI Cropping

We attempted to use a number of coronal offsets from the posterior edge of brain segmentations (Table 4.6), beginning with those which would place the ROI in the

vicinity of the cerebrospinal tract: This was also the approximate location of the slices selected by an expert in previous work using the CALSNIC-1 dataset [8]. For centre 1, it was discovered that the ROI which allowed for the greatest classification performance was located slightly anterior to the expected position surrounding the CST (see Figures 3.3 and 2.5). Nevertheless, there appears to be some overlap in the posterior areas of the ROI, while the anterior regions appear to include parts of the frontal lobes. Additionally, the ROI seems to overlap with the precentral gyrus, which is another motor region that appears to be impacted by the presence of ALS.

Table 4.6: Comparison of offsets from the brain posterior for the coronal slices. Note that feature selection was not applied due to the differences in ROIs. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Offset	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
0.42	2592	0.575	0.698	0.369	0.147	0.022	0.483
0.47	2592	0.643	0.747	0.471	0.140	0.020	0.578
0.52	2592	0.811	0.911	0.643	0.126	0.016	0.754
0.57	2592	0.702	0.847	0.460	0.144	0.021	0.596
0.62	2592	0.629	0.774	0.386	0.144	0.021	0.515

Similarly, we varied the number of coronal slices to determine the optimal size for the ROI. The minimum number of slices in any particular dimension must be at least 3, to account for the GO calculation. Depending on the neighbourhood size and CM calculation plane, the minimum number of slices increases. As seen in Table 4.7, using a large number of slices also resulted in decreased classification performance, suggesting that full brain classification without ROI selection is less effective. Hence, there seems to be an optimal length for the ROI that varies with the dataset.

Furthermore, we attempted to crop the ROI axially by adjusting the number of axial slices starting from the top of the brain mask (Table 4.8). Our results resembled those of the coronal slice experiments in that there is a certain maximum, above

Table 4.7: Comparison of coronal slice amounts. Note that feature selection was not applied due to the differences in ROIs. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Coronal Slices	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
11	2592	0.706	0.818	0.518	0.141	0.020	0.634
13	2592	0.748	0.860	0.562	0.140	0.019	0.680
15	2592	0.811	0.911	0.643	0.126	0.016	0.754
17	2592	0.765	0.866	0.598	0.136	0.018	0.707
19	2592	0.777	0.884	0.597	0.127	0.016	0.713

or below which there is a decrease in classification performance. Interestingly, the extension of the ROI beyond 18 slices caused a significant decrease in the F-score, which is correlated with the inclusion of a considerable portion of the corpus callosum across data samples in the centre.

Table 4.8: Comparison of axial slice amounts. Note that feature selection was not applied due to the differences in ROIs. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Axial Slices	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
14	2592	0.738	0.849	0.552	0.142	0.020	0.669
16	2592	0.801	0.894	0.646	0.127	0.016	0.750
18	2592	0.811	0.911	0.643	0.126	0.016	0.754
20	2592	0.740	0.854	0.551	0.135	0.018	0.670
22	2592	0.700	0.825	0.492	0.134	0.018	0.617

Scaling

The scaling of the image volume during pre-processing is a parameter also present in M-CoHOG. Nevertheless, instead of scaling individual slices, V-CoHOG uses cubic interpolation in order to scale the entire brain volume along 3 dimensions. A factor

of 0.33 was found to perform the best, which is the same magnitude used for scaling in M-CoHOG (Table 4.9). Using the image volume without scaling resulted in comparatively good specificity but very poor sensitivity and overall accuracy.

Table 4.9: Comparison of scaling factors. Note that feature selection was not applied due to the differences in ROIs. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Scaling	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
1	2592	0.577	0.516	0.679	0.155	0.024	0.587
0.5	2592	0.737	0.852	0.545	0.138	0.019	0.665
0.4	2592	0.716	0.811	0.557	0.136	0.018	0.660
0.33	2592	0.811	0.911	0.643	0.126	0.016	0.754
0.25	2592	0.572	0.703	0.353	0.144	0.021	0.470

Co-occurrence Matrix Calculation Parameters

While the previous parameters were utilized in the pre-processing step of V-CoHOG, the remaining parameters, which play a role prior to feature selection, impact the co-occurrence matrices. Both the neighbourhood size and number of orientation bins affect the number of features, while the gradient orientation type and co-occurrence matrix calculation plane do not. Only one direction was considered for the CM plane, resulting in 3 possible parameter settings, as opposed to 6. The maximum neighbourhood radius tested was 3 as values of 4 or greater resulted in a feature vector size that was too large for the logistic regression kernel machine base classifier.

The results for neighbourhood radii of sizes 2 and 3 were comparable (Table 4.10), with size 2 achieving a greater F-score while size 3 allowed for better accuracy. We chose to proceed with a value of 3 because it allowed for a greater feature size, increasing the selection pool of the feature selection step. Interestingly, a radius of size 1, which performed the best in the M-CoHOG method, yielded less optimal results, suggesting that V-CoHOG appears to be more effective with CoHOG features

that are more global in nature compared to the more localized features used for the Centre 1 dataset with M-CoHOG.

As was the case for M-CoHOG, 12 orientation bins allowed for the greatest classification performance (Table 4.11), with other bin numbers resulting in comparatively low specificities. Finally, the Axial-Coronal GO allowed for the greatest classifier performance with all three CM calculation planes (Table 4.12), although the Sagittal plane achieved the greatest accuracy overall.

Table 4.10: Comparison of neighbourhood size radii used for the CM calculation. Note that feature selection was not applied due to the differences in the feature vector sizes. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Radius	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
1	576	0.734	0.819	0.592	0.140	0.020	0.687
2	1440	0.807	0.897	0.658	0.130	0.017	0.759
3	2592	0.811	0.911	0.643	0.126	0.016	0.754

Table 4.11: Comparison of orientation bins for the calculation of the gradient orientations and co-occurrence matrices. Note that feature selection was not applied due to the differences in the feature vector sizes. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Bins	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
6	648	0.677	0.789	0.490	0.147	0.022	0.604
8	1152	0.708	0.871	0.436	0.146	0.021	0.581
10	1800	0.678	0.797	0.479	0.144	0.021	0.599
12	2592	0.811	0.911	0.643	0.126	0.016	0.754
14	3528	0.722	0.860	0.491	0.135	0.018	0.625

Feature Selection

Using our feature selection approach, we were able to reduce our feature vector size from 2592 to only 258, while improving both the accuracy and F-Score (Table 4.13).

Table 4.12: Comparison of gradient orientation types and the planes along which the co-occurrence matrix is calculated. A-S refers to Axial-Sagittal, A-C refers to Axial-Coronal, and S-C refers to Sagittal-Coronal. Note that feature selection was not applied due to the differences in GO and CM. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

GO Type	CM Plane	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
A-S	Coronal	2592	0.649	0.762	0.459	0.149	0.022	0.573
A-S	Sagittal	2592	0.638	0.774	0.412	0.146	0.021	0.538
A-S	Axial	2592	0.547	0.684	0.320	0.143	0.020	0.436
A-C	Coronal	2592	0.722	0.855	0.500	0.139	0.019	0.631
A-C	Sagittal	2592	0.811	0.911	0.643	0.126	0.016	0.754
A-C	Axial	2592	0.748	0.865	0.554	0.139	0.019	0.676
S-C	Coronal	2592	0.606	0.786	0.306	0.134	0.018	0.441
S-C	Sagittal	2592	0.677	0.827	0.428	0.136	0.019	0.564
S-C	Axial	2592	0.621	0.807	0.309	0.134	0.018	0.447

As is visible in our results, the classification performance was generally inversely proportional to the number of features until fewer than 258 features were selected. While feature numbers slightly larger than 258 provided slightly superior sensitivity scores, the specificity was significantly lower, resulting in inferior F-Scores.

Classifier Ablation Study

The results of our classifier ablation study indicate that the ensemble classifier is able to outperform the individual base classifiers in terms of both accuracy and F-Score (Table 4.14). Of the base classifiers, the two SVMs achieve the best performance while the KNN classifier performs the worst. While most of the base classifiers yielded accuracies and F-Scores within approximately 5% of the ensemble model, KNN results were significantly worse, with the exception of the sensitivity. Although the SVM and Logistic Regression Kernel Machine classifiers also achieve greater sensitivity scores than the ensemble model, all of the base classifiers have specificity scores below 0.79,

Table 4.13: Comparison of feature vector sizes used for classification based on ReliefF ranking. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
2592	0.811	0.911	0.643	0.126	0.016	0.754
2300	0.810	0.907	0.649	0.017	0.132	0.756
1896	0.811	0.921	0.628	0.127	0.016	0.747
1394	0.828	0.931	0.656	0.129	0.017	0.770
1043	0.826	0.920	0.668	0.123	0.015	0.774
776	0.851	0.937	0.706	0.120	0.014	0.805
559	0.853	0.926	0.732	0.117	0.014	0.817
396	0.858	0.936	0.727	0.114	0.013	0.819
258	0.860	0.889	0.812	0.118	0.014	0.848
154	0.821	0.826	0.812	0.130	0.017	0.819

which is reflected in their lower respective F-Scores.

Table 4.14: Ablation study of the ensemble classifier for Centre 1. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	258	0.860	0.889	0.812	0.118	0.014	0.848
Linear SVM	258	0.851	0.893	0.782	0.116	0.014	0.834
RBF SVM	258	0.850	0.894	0.778	0.117	0.014	0.832
KNN	258	0.690	0.948	0.260	0.104	0.011	0.408
FCNN	258	0.827	0.881	0.735	0.130	0.017	0.802
LogReg KM	258	0.846	0.907	0.744	0.118	0.014	0.817

Comparison with 3D CNNs

In order to compare our results to existing approaches, we used 3D implementations of ResNet10 [85], MobileNetv2 [86], ShuffleNetv2 [87] and DenseNet121 [88] as baselines.

The pre-processed ROIs of the dataset were padded with 0 values in order to create volumes of equal size in all 3 dimensions for the CNN methods after which they could be split into 5 folds. Therefore, the reported CNN results are the averages across the five folds where 4 folds were used for training and 1 for testing. In the case of the Centre 1 dataset, V-CoHOG outperformed all tested CNN approaches (Table 4.15) with the relatively complex DenseNet121 achieving the worst results. As CNN methods normally rely on an abundance of training data in order to achieve excellent classification performance, comparatively poor performance seen in our results is not surprising, considering the small size of the Centre 1 dataset.

Table 4.15: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the Centre 1 dataset. The CNN approaches achieve comparably low accuracies and F-Scores, likely due to the size of the dataset. The accuracy (Acc), sensitivity (Sens), and specificity (Spec) have been abbreviated.

Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.860	0.889	0.812	0.848
ResNet10	0.632	0.610	0.667	0.627
MobileNetv2	0.632	0.700	0.533	0.374
ShuffleNetv2	0.632	0.680	0.533	0.335
DenseNet121	0.604	0.650	0.533	0.240

Feature Mapping

Using the 258 features from our selection step which yielded the best classification performance, we create segmentation maps of the origin voxels in the GO volumes of select patients and controls. We then overlaid these segmentation maps onto their original imaging volumes (Figure 4.2). As the segmentation maps are based on the GO volumes calculated using the pre-processed imaging data, we estimate the overlay accuracy to be within slightly over 3mm due to the effects of 0.33 factor scaling and the Sobel operation on the input volume.

While 2 of these 4 samples were usually mis-classified, all four appear to share

a number of segmented areas for multiple features, as marked by the brightest regions in the segmentation maps. These areas are most visible in the coronal slices (right column) and appear to be positioned slightly above the corpus callosum, in the approximate vicinity of the CST's anterior boundary. Therefore, it appears that the origins of some of the features generated using the V-CoHOG method localize to comparable regions of the brain, even without the use of normalization or registration in the pre-processing stage.

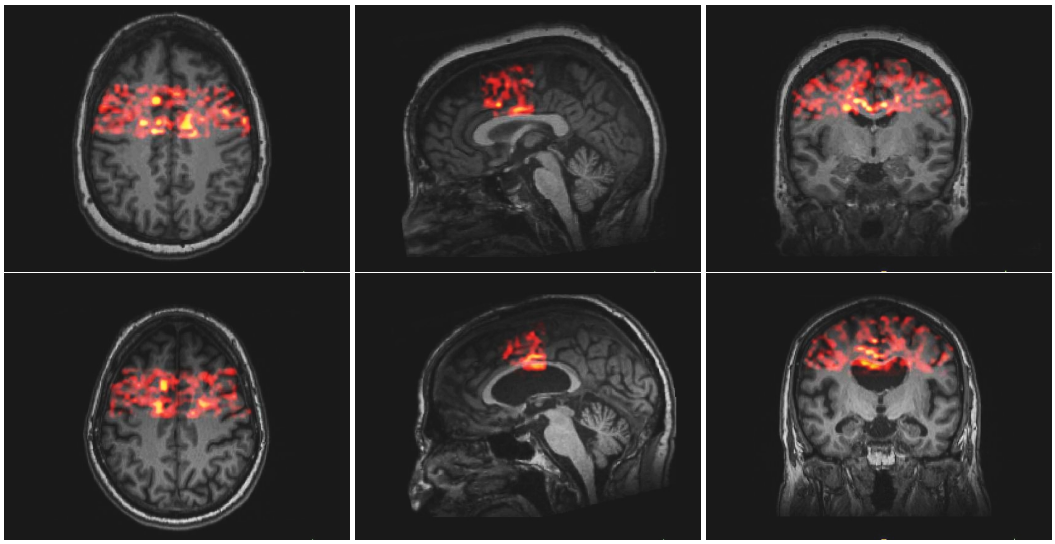
4.3.2 Centre 2

Centre 2 was the second largest in terms of dataset size and involved the use of a GE MRI scanner for image acquisition as opposed to Centre 1's Siemens. While most of the optimal parameter values were shared with Centre 1, a coronal offset of 0.56 for the ROI was found to allow for better performance than Centre 1's 0.52 (Table 4.4). Additionally, only 174 features were required to achieve the optimal results.

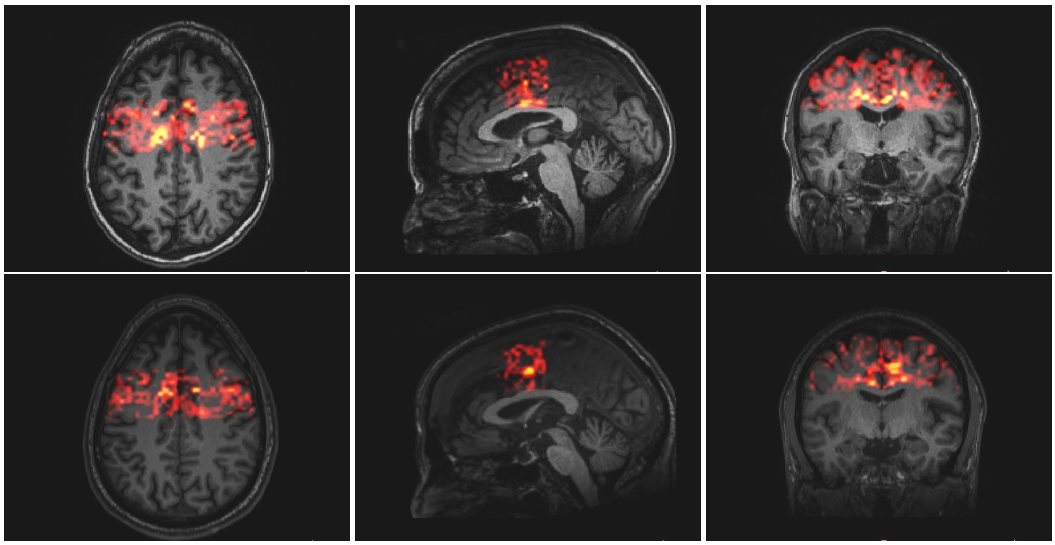
The achieved accuracy of 0.792 (Table 4.5) is enough to surpass M-CoHOG's 0.766 for the same dataset. However, due to the low specificity, M-CoHOG's F-Score of 0.763 is superior to the proposed method's 0.730. The ablation result for centre 2 (Table 4.16) is generally similar to that of centre 1. However, the FCNN base classifiers appear to be significantly worse than what was observed previously. Nevertheless, KNN results remain the lowest while the ensemble classifier once again achieves the top performance.

Comparison with 3D CNNs

CNN results for the Centre 2 dataset were similar to those of Centre 1 (4.17), with V-CoHOG again achieving superior classification performance to the 4 deep learning methods. However, both ResNet10 and MobileNetv2 achieve superior specificity to our approach, although it is not enough to provide a significant improvement to their respective F-Scores.



(a) Patients



(b) Controls

Figure 4.2: Heatmaps of the selected 258 features overlaid onto a selection of patients and controls from Centre 1 which were predominantly classified either correctly or incorrectly. Areas with greater intensity correspond to regions which acted as the origin of a greater number of features. The top 2 rows belong to ALS patient data samples while the bottom 2 rows were controls. Within each group, the top row corresponds to a data sample that was correctly classified an overwhelming number of times while the bottom was usually misclassified. Note that the second row of the patient section corresponds to a data sample that was determined to be an ALS mimic rather than a true positive and later removed as part of a quality review.

Table 4.16: Ablation study of the ensemble classifier for Centre 2. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

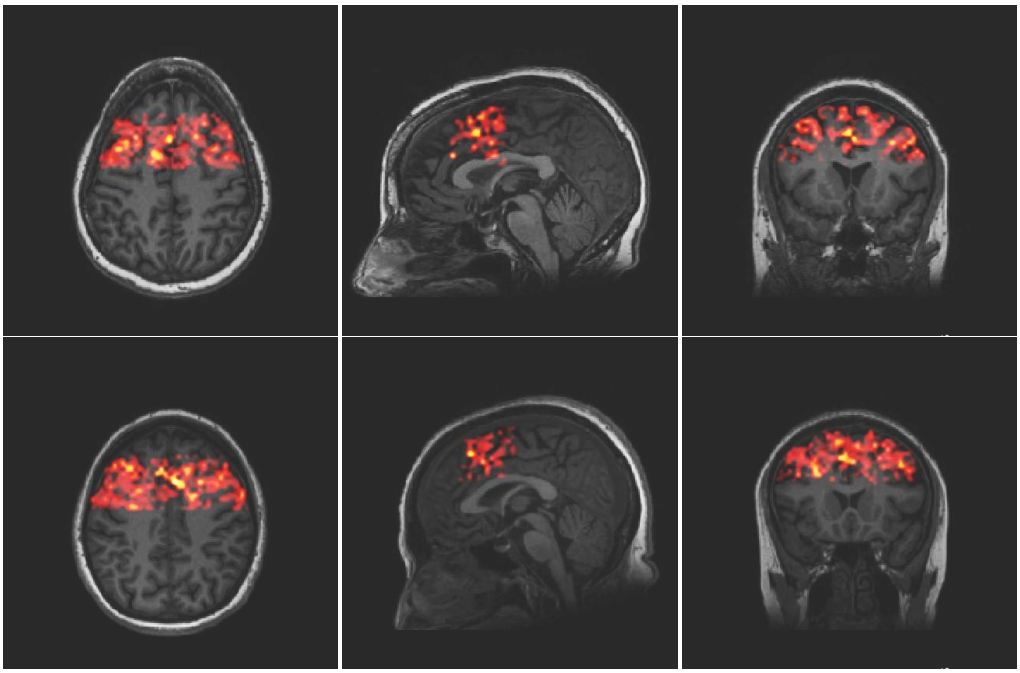
Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	174	0.792	0.874	0.627	0.147	0.022	0.730
Linear SVM	174	0.781	0.896	0.551	0.145	0.021	0.682
RBF SVM	174	0.782	0.897	0.552	0.145	0.021	0.683
KNN	174	0.665	0.998	0.000	0.015	0.000	0.000
FCNN	174	0.684	0.941	0.170	0.103	0.011	0.288
LogReg KM	174	0.766	0.898	0.501	0.139	0.019	0.643

Table 4.17: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the Centre 2 dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

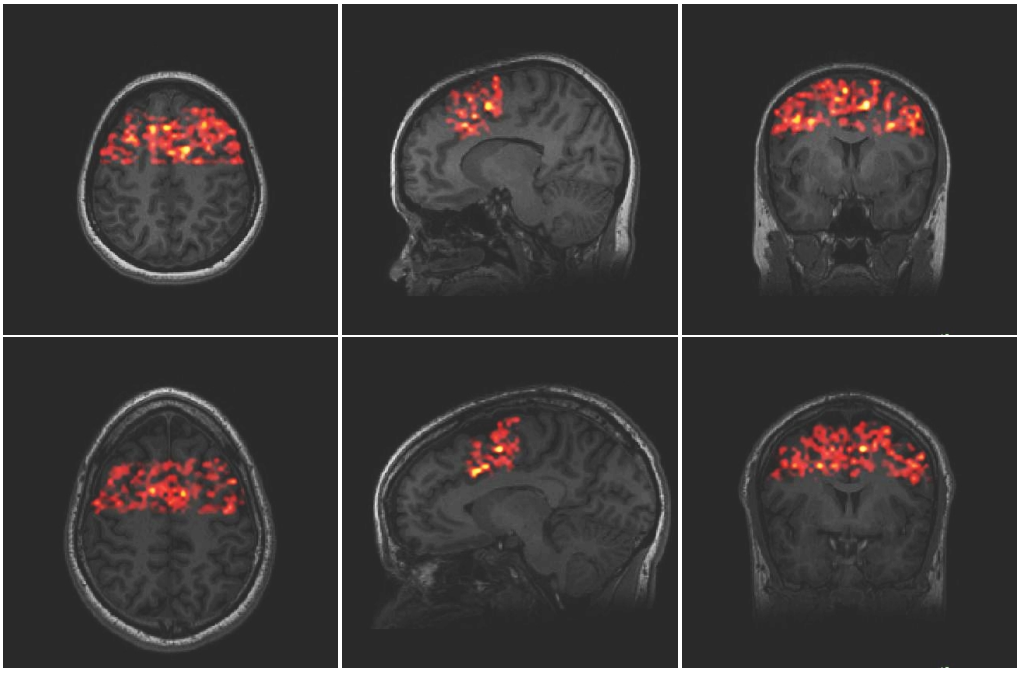
Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.792	0.874	0.627	0.730
ResNet10	0.675	0.590	0.833	0.666
MobileNetv2	0.442	0.230	0.833	0.300
ShuffleNetv2	0.500	0.450	0.600	0.080
DenseNet121	0.483	0.400	0.600	0.000

Feature Mapping

The feature segmentation masks for Centre 2 (Figure 4.3) appear more heterogeneous than those of Centre 1, particularly when comparing patients (top 2 rows) with controls (bottom 2 rows). This observation may be partially explained by the significant difference in sensitivity and specificity obtained for this dataset (Table 4.5). Due to the greater ROI offset from the anterior, the features now also seem to be located mostly outside of CST regions.



(a) Patients



(b) Controls

Figure 4.3: Heatmaps of the selected 174 features overlaid onto a selection of patients and controls from Centre 2 which were predominantly classified either correctly or incorrectly. Areas with greater intensity correspond to regions which acted as the origin of a greater number of features. The top 2 rows belong to ALS patient data samples while the bottom 2 rows were controls. Within each group, the top row corresponds to a data sample that was correctly classified an overwhelming number of times while the bottom was usually misclassified.

4.3.3 Centre 3

Even though Centre 3 shared it's MRI system vendor with Centre 2, we found the parameters used for Centre 1 to be optimal (Table 4.4). Using a selection of 186 features, we obtained an accuracy of 0.791 (Table 4.5), once again outperforming the M-CoHOG method's 0.732. Additionally, both the KNN and FCNN base classifiers were not able to yield any promising results (Table 4.18), similar to what was seen for Centre 2. Overall, the results remain in line with those of the previous two centres, with a sensitivity that is significantly greater than the specificity, possibly impacted by the imbalance in the number of patients and controls in the training set.

Table 4.18: Ablation study of the ensemble classifier for Centre 3. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated. Due to the comparatively low classification performance of the KNN and FCNN base classifiers, another ablation study was performed that did not include these two classifiers in Section 4.3.4.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	186	0.791	0.868	0.715	0.209	0.044	0.784
Linear SVM	186	0.788	0.867	0.709	0.207	0.043	0.780
RBF SVM	186	0.788	0.867	0.710	0.207	0.043	0.780
KNN	186	0.500	1.000	0.000	0.000	0.000	0.000
FCNN	186	0.500	1.000	0.001	0.008	0.000	0.001
LogReg KM	186	0.773	0.834	0.712	0.201	0.041	0.768

Comparison with 3D CNNs

Once again, V-CoHOG outperformed the 3D CNN methods tested (Table 4.19), although there was an increase in the accuracy achieved by some of the deep learning approaches. Nevertheless, the F-Score remained significantly below that of our method.

Table 4.19: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the Centre 3 dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.791	0.868	0.715	0.784
ResNet10	0.737	0.733	0.700	0.581
MobileNetv2	0.520	0.800	0.200	0.000
ShuffleNetv2	0.730	0.600	0.900	0.633
DenseNet121	0.520	0.533	0.500	0.114

4.3.4 Centre 4

The last of the CALSNIC-1 datasets, Centre 4 also shared most optimal parameters with Centre 1 with the exception of a slight difference in the coronal offset (Table 4.4). When provided to the ensemble classifier, the 164 features selected allowed for the highest single accuracy among the centres at 0.870 (Table 4.5), significantly outperforming M-CoHOG's result of 0.716. While the KNN base classifier was again unable to achieve a meaningful result (Table 4.20), the SVM classifier with a linear kernel obtained an F-Score that was slightly higher than that of the ensemble classifier, demonstrating that the ensemble classifier may be slightly outperformed by particular base classifiers in certain cases.

Comparison with 3D CNNs

Centre 4 saw the best performance of a 3D CNN classifier among the single-centre datasets, with ResNet10 and MobileNetv2 achieving specificities of 1.000 and ResNet10 achieving an F-Score of 0.787. While ShuffleNetv2 managed to obtain an accuracy of 0.813, the F-Score remained lacking. Therefore, our method still attained the greatest accuracy and F-Score.

Table 4.20: Ablation study of the ensemble classifier for Centre 4. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	164	0.870	0.895	0.792	0.167	0.028	0.841
Linear SVM	164	0.846	0.847	0.845	0.173	0.030	0.846
RBF SVM	164	0.842	0.842	0.842	0.173	0.030	0.842
KNN	164	0.750	1.000	0.000	0.000	0.000	0.000
FCNN	164	0.857	0.892	0.754	0.149	0.022	0.817
LogReg KM	164	0.853	0.872	0.796	0.166	0.028	0.832

Table 4.21: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the Centre 4 dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.870	0.895	0.792	0.841
ResNet10	0.787	0.667	1.000	0.787
MobileNetv2	0.740	0.600	1.000	0.660
ShuffleNetv2	0.813	0.867	0.700	0.633
DenseNet121	0.557	0.800	0.200	0.000

Removal of KNN and FCNN Base Classifiers

In the single-centre experiments, the KNN base classifier tended to perform significantly worse than the other classifiers. The FCNN classifier yielded lower classification results than the SVM and Logistic Regression Kernel Machine classifiers, especially for the Centre 2 (Table 4.16) and 3 (Table 4.18) datasets. We investigated the effect of removing the KNN and FCNN classifiers from the ensemble model (Table 4.22) but did not observe significant improvements to the classification results.

Table 4.22: Results of the ensemble classifiers for Centres 1-4. Ensemble Size 5 results use all 5 base classifiers in the ensemble model while the Ensemble Size 3 results were generated using an ensemble model with SVM (linear kernel), Logistic Regression Kernel Machine and another SVM (RBF kernel) base classifiers. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Centre	Ensemble Size	Acc	Sens	Spec	Std Dev	Var	F-Score
1	5	0.860	0.889	0.812	0.118	0.014	0.848
1	3	0.848	0.874	0.804	0.120	0.014	0.837
2	5	0.792	0.874	0.627	0.147	0.022	0.730
2	3	0.790	0.884	0.603	0.145	0.021	0.717
3	5	0.791	0.868	0.715	0.209	0.044	0.784
3	3	0.787	0.876	0.698	0.202	0.041	0.777
4	5	0.870	0.895	0.792	0.167	0.028	0.841
4	3	0.851	0.869	0.799	0.175	0.030	0.832

4.4 Multi Centre Datasets

4.4.1 Centre 1 and 4

In order to better understand the impact of multi-centre datasets on the classification accuracy, we combined the Centre 1 and 4 datasets, as they share MRI scanner vendors and achieved the best classification results (Table 4.5), and performed a feature selection on the unified dataset. For all the CALSNIC-1 multi-centre experiments, the 80-20 train-test split was applied to the data on a per centre basis. We observed a considerable decrease in classification performance, specifically when comparing F-Scores to those obtained in the two centres separately (Table 4.23). This result, which bares the hallmarks typical of those for multi-centre classification studies, could suggest that factors which are not consistent due to the use of the scanner vendor, such as imaging technique, are impacting the texture of the image volumes. Hence, even with the same MRI system vendor and the CALSNIC directives to enhance comparability, combining data from centres 1 and 4 results in a significant decrease in classification

performance.

Table 4.23: Ablation study of the ensemble classifier for the Centre 1 & 4 combined dataset. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	195	0.765	0.772	0.752	0.108	0.012	0.762
Linear SVM	195	0.784	0.835	0.683	0.102	0.011	0.751
RBF SVM	195	0.772	0.827	0.662	0.104	0.011	0.735
KNN	195	0.729	0.872	0.443	0.107	0.011	0.587
FCNN	195	0.797	0.835	0.720	0.104	0.011	0.773
LogReg KM	195	0.811	0.873	0.686	0.099	0.010	0.768

4.4.2 Centre 2 and 3

Similarly to the Centre 1 and 4 dataset, we combined the data from Centre 2 and 3 which both used scanners manufactured by GE. Once again there is a decrease in classification performance, with the various base classifiers achieving relatively low specificity scores (Table 4.24). Given our results for these combined datasets, it appears that the comparability enhancing directives laid out by CALSNIC-1 were not sufficient to prevent a number of the problems introduced by multi-centre datasets.

4.4.3 CALSNIC-1

The combined CALSNIC-1 dataset encompassed all 4 centres and a parameter search revealed that the optimal parameters were different for both the pre-processing as well as the CM calculation steps (Table 4.4). While the optimal neighbourhood radius and number of orientation bins seemed to be consistent across single centres and the joint dataset, the GO type and CM plane values found to be best for centre 1 were only optimal for the other individual centres but not for the combined CALSNIC-1 dataset. Furthermore, in order to help keep the ROI more consistent across centres,

Table 4.24: Ablation study of the ensemble classifier for the Centre 2 & 3 combined dataset. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	89	0.719	0.758	0.660	0.126	0.016	0.706
Linear SVM	89	0.750	0.868	0.575	0.117	0.014	0.691
RBF SVM	89	0.751	0.868	0.575	0.014	0.014	0.692
KNN	89	0.729	0.930	0.427	0.099	0.010	0.585
FCNN	89	0.684	0.940	0.301	0.097	0.009	0.456
LogReg KM	89	0.759	0.894	0.556	0.116	0.013	0.685

the number of axial slices was no longer determined starting from the top of the brain volume mask, but rather from the computed centroid of the mask volume. The optimal number of selected features was determined to be 268 and when supplied to the ensemble classifier resulted in an accuracy of 0.770 (Table 4.25). This result slightly outperforms the 0.745 accuracy score achieved by M-CoHOG without centre-wise weighting (Table 4.5). Unlike the majority of the single-centre experiments, the FCNN base classifier outperformed the Logistic Regression Kernel Machine. All base classifiers achieved comparable sensitivity results while the significantly lower specificity scores accounted for the differences between accuracy and F-Score metrics.

Table 4.25: Ablation study of ensemble classifier results for the CALSNIC-1 dataset. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	268	0.770	0.820	0.684	0.075	0.006	0.746
Linear SVM	268	0.768	0.827	0.666	0.075	0.006	0.738
RBF SVM	268	0.768	0.826	0.665	0.075	0.006	0.737
KNN	268	0.648	0.882	0.237	0.073	0.005	0.374
FCNN	268	0.760	0.837	0.626	0.081	0.007	0.717
LogReg KM	268	0.740	0.875	0.505	0.080	0.006	0.640

Comparison with 3D CNNs

Compared to the 4 3D CNN methods, our method was able to achieve the highest accuracy and F-Score results (Table 4.26). Even though the combined CALSNIC-1 dataset was considerably larger than those of the individual centres, the difference did not appear significant enough to allow the CNN methods to achieve comparable accuracy or F-Score results to those of V-CoHOG.

Table 4.26: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the combined CALSNIC-1 dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.770	0.820	0.684	0.746
ResNet10	0.641	0.683	0.567	0.593
MobileNetv2	0.632	0.670	0.572	0.616
ShuffleNetv2	0.590	0.476	0.783	0.527
DenseNet121	0.481	0.400	0.600	0.000

Feature Mapping

While the feature maps generated for the CALSNIC-1 dataset are also less uniform than those of from Centre 1 (Figure 4.4), there appear to be some regions with greater feature intensity that are consistent across the samples. Because the image volumes were derived from different centres, some of the slices shown have different brightness levels (see row 2) compared with the others.

We also provide a comparison of the same slices from 2 patients with their single centre and multi-centre maps placed next to each other (Figure 4.5). Overall the ROI for the combined dataset falls roughly over the area expected to contain the CST, hence, a close examination of the areas where the posterior features of the single centre map coincide with those also present in the CALSNIC-1 feature map could reveal more information regarding potential biomarkers. For example, as can

be seen in the coronal slice images (right column), the region slightly above the corpus callosum contains overlapping multi-feature areas for the patient represented in the top 2 rows.

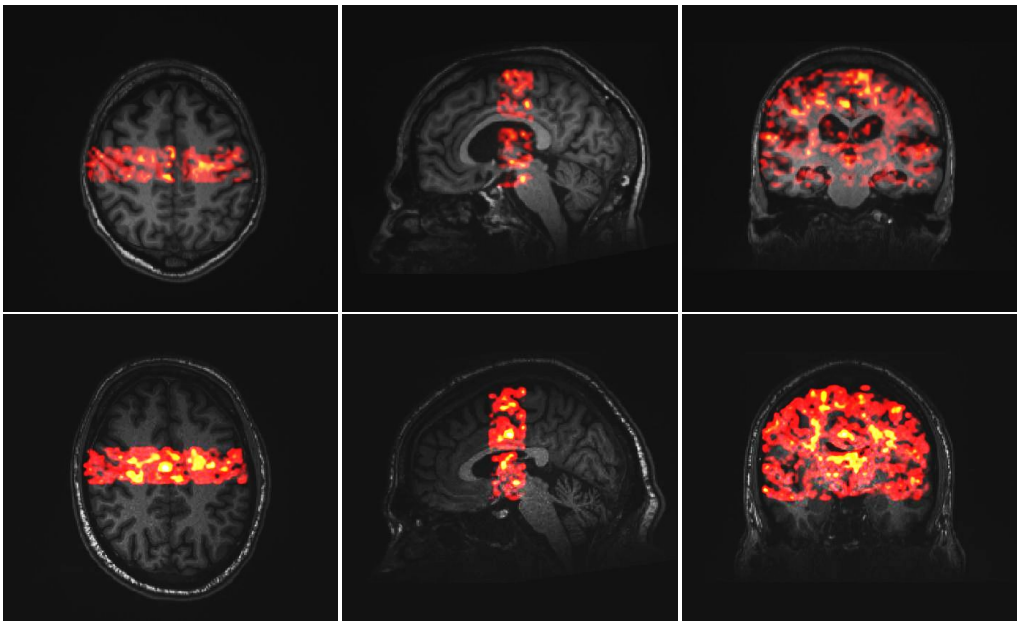
4.4.4 Post Data Review

After we had generated our results for the CALSNIC-1 dataset, a quality review of the data revealed that some of the patient samples in the four centres had been incorrectly labelled and belonged to subjects with conditions that act as ALS mimics, such as Multiple Sclerosis (MS). Feature maps from one such patient are present in the second row of the feature maps for Centre 1 (Figure 4.2). This may explain the poor classification performance for this particular subject which was classified as a patient only 22 out of 244 runs in the Centre 1 experiment. However, the same subject was classified as a patient 183 out of 199 runs in experiments with the combined CALSNIC-1 dataset. Hence, the multi-centre features could be decreasing the ensemble model's ability to differentiate between ALS patients and subjects with a mimic condition.

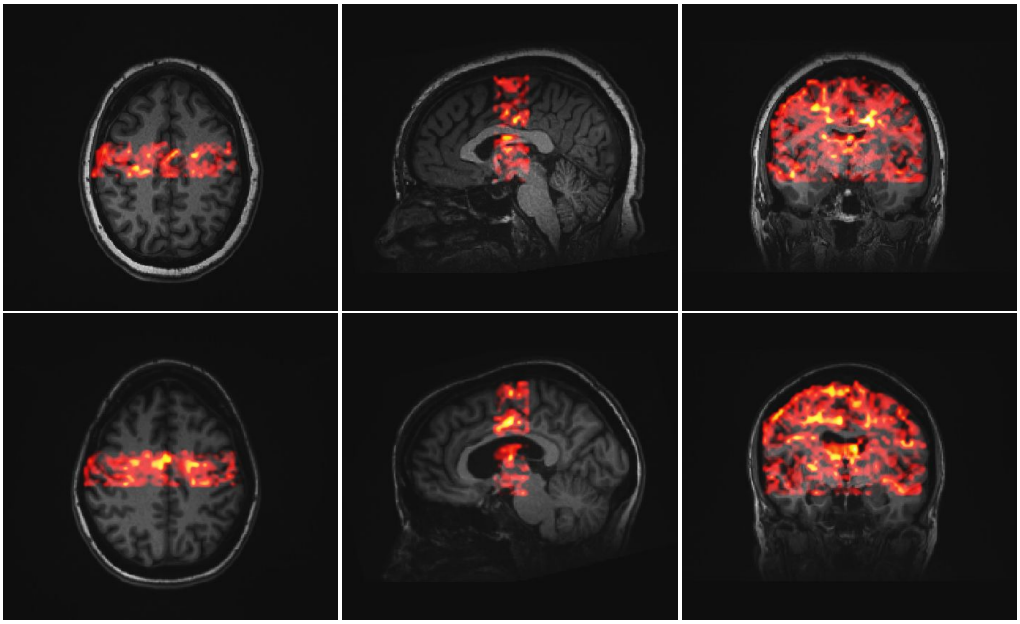
Nevertheless, removing patient samples which had been identified as problematic, which reduced the total number of samples in the CALSNIC-1 dataset from 111 to 96 (Table 4.27), did not appear to have any significant effects on most of the classification results (Table 4.28). In a number of cases, the accuracy and/or F-Score slightly decreased. The exception was the combined CALSNIC-1 dataset, which saw an increase in specificity to 0.738, accuracy to 0.781 and F-Score to 0.773, suggesting that the ALS mimic datasamples in the CALSNIC-1 dataset are contributing to the ensemble model's difficulty identifying control samples.

4.5 CALSNIC-2

CALSNIC-2 is a dataset which is still undergoing growth and has not seen use in ALS classification methods before. With the inclusion of 225 subjects, CALSNIC-2

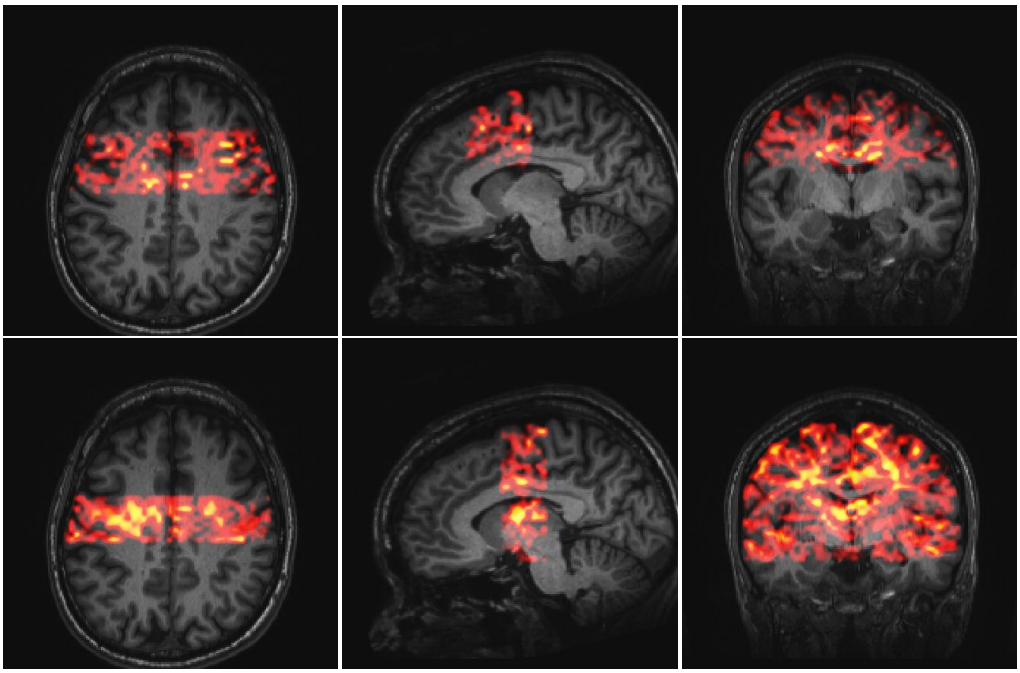


(a) Patients

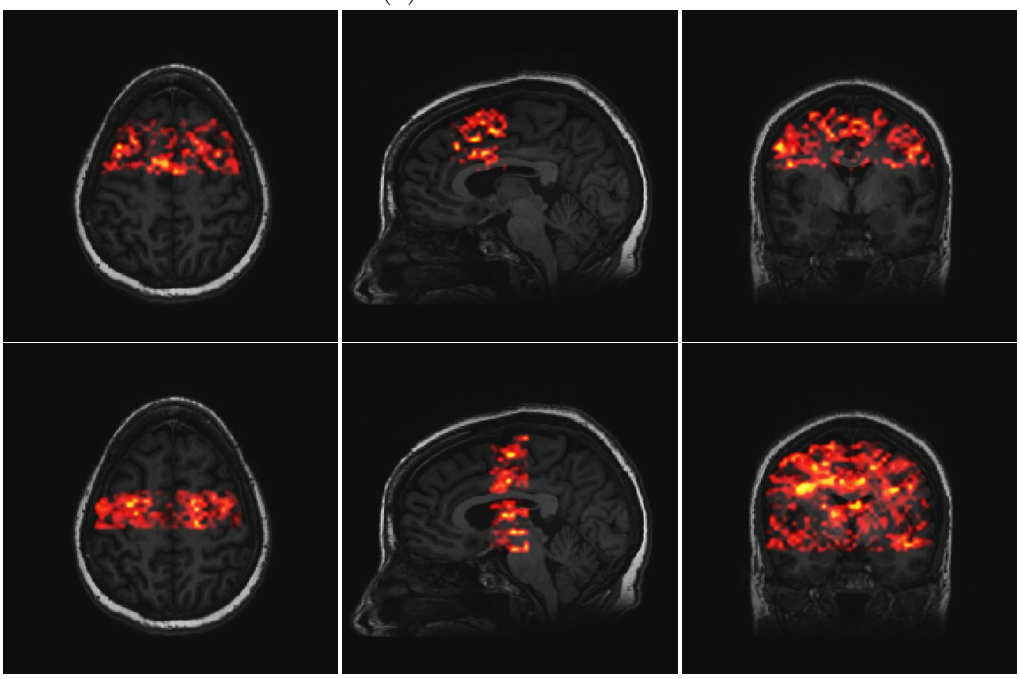


(b) Controls

Figure 4.4: Heatmaps of the selected 268 features overlaid onto a selection of patients and controls from the combined CALSNIC-1 dataset which were predominantly classified either correctly or incorrectly. Areas with greater intensity correspond to regions which acted as the origin of a greater number of features. The top 2 rows belong to ALS patient data samples while the bottom 2 rows were controls. Within each group, the top row corresponds to a data sample that was correctly classified an overwhelming number of times while the bottom was usually misclassified.



(a) Centre 1 Patient



(b) Centre 2 Patient

Figure 4.5: Heatmap comparison of a patient from Centre 1 (top 2 rows) and Centre 2 (bottom 2 rows). The first row in each section corresponds to the selected features used when classifying only within the same centre while the second row corresponds to the features used for classification in the CALSNIC-1 multicentre dataset.

Table 4.27: Overview of the number of patients and controls in the 4 CALSNIC-1 centres after a quality review. All control samples were retained from the initial round of experiments.

Centre	Edmonton (1)	Toronto (2)	Calgary (3)	Montreal (4)
Patients	19	16	8	11
Controls	15	12	8	7

Table 4.28: Ensemble classifier results for the CALSNIC-1 centres after a quality review. The accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Centre	Acc	Sens	Spec	Std Dev	Var	F-Score
1	0.848	0.930	0.739	0.125	0.016	0.824
2	0.738	0.790	0.659	0.176	0.031	0.719
3	0.824	0.884	0.763	0.190	0.036	0.819
4	0.859	0.883	0.812	0.203	0.041	0.846
1+4	0.796	0.880	0.671	0.103	0.011	0.761
2+3	0.718	0.725	0.709	0.118	0.014	0.717
All	0.781	0.812	0.738	0.089	0.008	0.773

is comparatively large given the sizes of other imaging datasets in the ALS domain. While the nature of the dataset shares numerous traits with CALSNIC-1 and includes even more centres, a parameter search yielded slight differences when comparing the optimal parameter values (Table 4.4). Additionally, a relatively small feature set of size 71 allowed for the best classification performance (Table 4.29) with the ensemble classifier achieving better accuracy and F-Score metrics than all of the base classifiers.

4.5.1 Comparison with 3D CNNs

Compared to the four 3D CNN approaches tested, V-CoHOG achieved both the highest accuracy and F-Score results (Table 4.30). Possibly due to the increased size of the dataset, ResNet10 was able to achieve results which were superior to those obtained with the CALSNIC-1 dataset, even if the metrics were not great enough to

Table 4.29: Ablation study of ensemble classifier results for the CALSNIC-2 dataset. The number of features (Feat), accuracy (Acc), sensitivity (Sens), specificity (Spec), standard deviation (Std Dev) and variance (Var) have been abbreviated.

Classifier	Feat	Acc	Sens	Spec	Std Dev	Var	F-Score
Ensemble	71	0.752	0.733	0.771	0.058	0.003	0.752
Linear SVM	71	0.743	0.699	0.790	0.058	0.003	0.741
RBF SVM	71	0.743	0.699	0.790	0.058	0.003	0.741
KNN	71	0.702	0.569	0.841	0.063	0.004	0.679
FCNN	71	0.746	0.718	0.774	0.003	0.059	0.745
LogReg KM	71	0.747	0.713	0.782	0.061	0.004	0.746

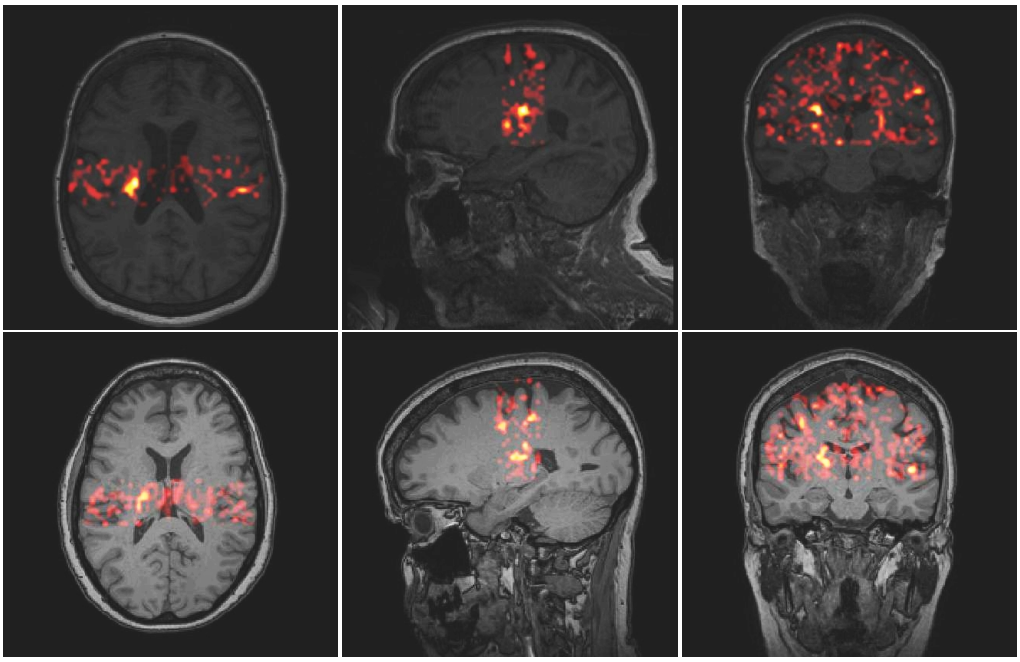
surpass our method.

Table 4.30: Results of the V-CoHOG ensemble classifier as well as those of the 4 3D CNN methods for the CALSNIC-2 dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

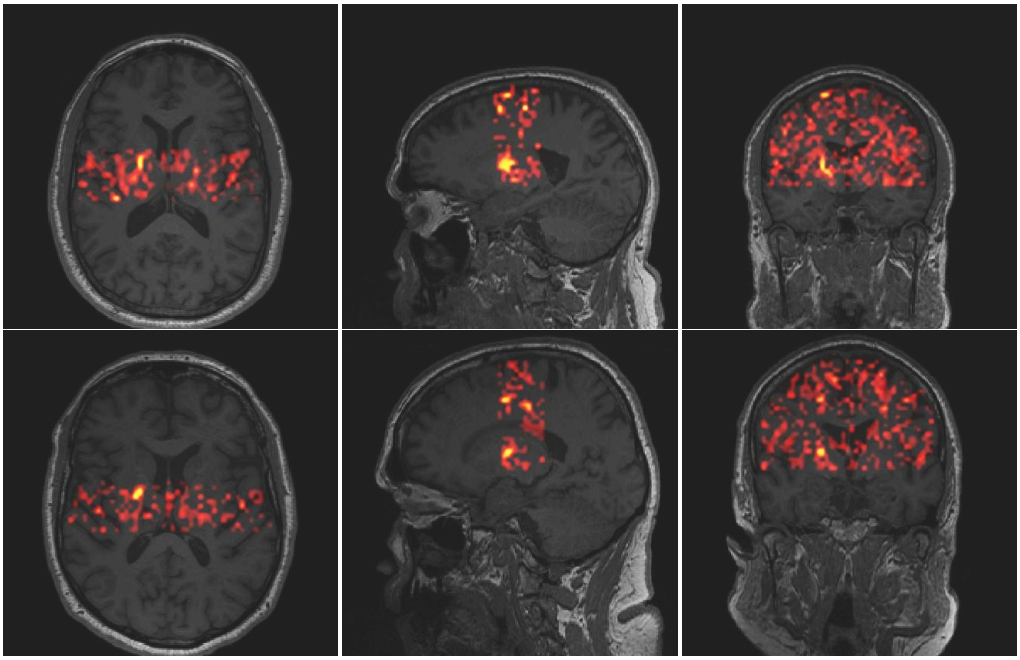
Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.752	0.733	0.771	0.752
ResNet10	0.700	0.698	0.702	0.685
MobileNetv2	0.615	0.558	0.669	0.589
ShuffleNetv2	0.627	0.624	0.631	0.613
DenseNet121	0.532	0.508	0.557	0.398

4.5.2 Feature Mapping

Interestingly, the segmentation maps for the CALSNIC-2 dataset appear more homogenous than those observed for the CALSNIC-1 dataset (Figure 4.6). Notably, there appears to be an origin region of multiple selected features located to slightly below the right side of the corpus callosum (this region appears on the left instead of right in the coronal slices). Once again, a number of features appear to originate in regions that could correspond to areas of the CST.



(a) Patients



(b) Controls

Figure 4.6: Heatmaps of the selected 71 features overlaid onto a selection of patients and controls from multicentre CALSNIC-2 dataset which were predominantly classified either correctly or incorrectly. Areas with greater intensity correspond to regions which acted as the origin of a greater number of features. The top 2 rows belong to ALS patient data samples while the bottom 2 rows were controls. Within each group, the top row corresponds to a data sample that was correctly classified an overwhelming number of times while the bottom was usually misclassified.

4.6 Application to Alzheimer's Disease

Another neurodegenerative affliction that is diagnosed using a combination of clinical presentation and biomarkers is Alzheimer's Disease (AD), which is responsible for more dementia cases around the world than any other cause [89]. From 2004 to 2011, AD was specified as the cause of 80,868 deaths in Canada [90]. In order to demonstrate that our method was applicable to diseases other than ALS, we chose a subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [91] to evaluate our methods performance when tasked with classifying between AD patients and healthy controls.

In this section, we present the results of a preliminary parameter search and classification using V-CoHOG in conjunction with the ADNI dataset. We selected a total of 172 AD patients and 229 healthy controls to use for the evaluation of our method and maintained the same evaluation criteria and approach as was used for the ALS datasets. The parameter search results suggest that most values used for the ALS datasets were not necessarily well suited for AD, with the exception of the quantization bin size (Table 4.31). Another important difference appears to be the importance of a comparatively large feature number of 1358, rather than the few hundred features which generally worked best for ALS. We use a threshold $tau = 3$ (Algorithm 1) to calculate the feature vector and $K = 13$ for feature selection with ReliefF using $K = 13$ (Section 3.3.2).

4.6.1 Comparison with 3D CNNs

When compared with ResNet10, which yielded the best overall CNN results for ALS, our method achieves a comparable accuracy and F-Score. Of note is that the sensitivity and specificity scores of our method and ResNet10 appear to be divergent, while the accuracy and F-Score appear to lie between the two disparate metrics. Hence, our method appears to be adaptable to a variety of volume classification tasks present in

Table 4.31: Parameters found to yield the optimal classification results for the ADNI dataset after a preliminary parameter search.

Parameter	Value
Scaling	0.5
Coronal Offset	0.42
Coronal Slices	26
Axial Slices	Centroid + 10
Radius	4
Bins	12
GO Type	Sagittal-Coronal
CM Calc Plane	Axial
Features	1358

Table 4.32: Results of the V-CoHOG ensemble classifier as well as those of ResNet10 for the ADNI dataset. The accuracy (Acc), sensitivity (Sens) and specificity (Spec) have been abbreviated.

Classifier	Acc	Sens	Spec	F-Score
V-CoHOG	0.763	0.708	0.805	0.753
ResNet10	0.758	0.849	0.690	0.757

the medical imaging domain.

Chapter 5

Conclusion

5.1 Summary

In this chapter, we conclude our thesis by providing a brief overview of the background, methodology and results discussed previously, and how these concepts can contribute to the field of texture-based volume classification in medical imaging. Specifically, we will address the runtime performance of CUDA-Accelerated M-CoHOG as well as the classification performance and observations of V-CoHOG when applied to the CALSNIC and ADNI MRI datasets. We will also present some possible directions for future work which builds upon our proposed method and findings.

First, in Chapter 2, we discussed the use of image textures in order to generate features which can be applied to statistic or classification tasks. One of the earliest texture feature approaches, Gray-Level Co-occurrence Matrices, calculated a variety of statistical features based on pixel intensities, or gray levels, and was followed by methods such as Gray-Level Run Lengths, wavelets and Scale Invariant Feature Transform which expanded the domain of invariant texture analysis. The development of Histograms of Oriented Gradients allowed for even greater classification performance than the previous methods in a number of tasks; with the development of various modifications and additions, such as Co-occurrence Histograms of Oriented Gradients, expanding upon the original gradient orientation and histogram computation approach. While CoHOG mainly introduced the concept of calculating the

CMs based on orientation pairs, most recent applications of HOG and CoHOG have involved further additions such as the use of multiple input image resolutions as well as adaptations to three dimensional imaging data.

Next, we introduced the concept of hardware accelerated image processing and discussed the advantages of using parallel computing in order to analyze digital imaging data. Notably, we looked at examples of HOG adaptations partially implemented to run on GPUs, and the runtime performance benefits of such approaches. The discussion of image processing was then directed towards the medical imaging domain, where fundamental principles behind magnetic resonance imaging were introduced. A number of limitations continue to pursue MRI datasets due to factors such as technique consistency, normalization methods, and relatively small dataset sizes. Many of these factors are further compounded in multi-centre datasets which attempt to address the size limitation. Furthermore, post-processing techniques applied to MRI data, while circumstantially beneficial, can hinder the performance of downstream tasks, such as classification.

Amyotrophic Lateral Sclerosis, a neurodegenerative disease with only rare cases of symptom reversion, was the focus of the majority of datasets we applied. After discussing the neuroanatomical background of the condition, we examined previous work on imaging biomarkers for ALS and how they may improve patient survival time via more rapid classification allowing for earlier treatment. Nevertheless, previously existing ALS MRI datasets involved the same limitations mentioned previously, hence, initiatives such as the Canadian ALS Neuroimaging Consortium were created in order to improve the intra-comparability of multi-centre imaging datasets while achieving an increased dataset size not attainable by a single location. Finally, we concluded the chapter by introducing Modified Co-occurrence Histograms of Oriented Gradients which was able to achieve state-of-the-art classification results on a CALSNIC-1 dataset.

Chapter 3 began by detailing our CUDA implementation of the costliest sections of

the M-CoHOG method. We then proceeded to introduce our proposed extension of M-CoHOG, Volumetric Co-occurence Histograms of Oriented Gradients, and discussed the changes made to the CUDA utilization when compared with the original method. Initially, we pre-processed the data by creating brain volume segmentation masks with Brain Extraction Tool which could then be cropped and scaled for ROI selection. The feature generation involved the computation of 3 gradient orientation volumes, one of which was then selected for calculation of the co-occurrence matrices using a neighbourhood size with a given radius. After the normalization and concatenation of the feature vector, we applied feature selection via ReliefF in order to choose a subset of features to be used for classification with our ensemble model classifier.

In Chapter 4 we discussed our results, beginning with an evaluation of the run-time performance of CUDA-accelerated M-CoHOG on the Centre 1 ALS and DIV2K datasets. While the speedup achieved was rather marginal when using small images, generation of features could be performed up to 19.9× as quickly for the DIV2K data subset. Thereafter, we examined the classification performance of our method on the ALS datasets CALSNIC-1 and CALSNIC-2, where our method achieved superior accuracy and F-scores to 4 3D CNN methods. Additionally, our method allowed for state-of-the-art classification performance on the CALSNIC-1 single-centre datasets when compared to M-CoHOG. Additionally, our method's accuracy of 0.770 was able to surpass the comparable 3D CNN approaches tested as well as the state-of-the-art for the combined and unweighted CALSNIC-1 dataset without relying on expert segmented ROI masks. For CALSNIC-2, V-CoHOG also achieved the best performance with an accuracy of 0.752, which could be considered a baseline result for future classification studies using the still expanding CALSNIC-2 dataset. Finally, applying our method to the ADNI dataset allowed us to achieve comparable performance to ResNet10 after a preliminary parameter search, with an accuracy of 0.763. Unfortunately, we were not able to find a fixed set of ROI cropping parameters which yielded superior results for all ALS datasets, as even slight changes to the ROI could result

in significant changes to classification performance (see 4.3.1). Nevertheless we observed a trend in that ROIs which were located more anteriorly performed better in the single-centre datasets, while ROIs more closely aligned with the CST allowed for better multi-centre metrics.

In addition to patient classification, we also generated segmentation maps for the selected feature sets, indicating their place of origin in the original imaging volumes. These maps displayed a number of segmentation clusters which appeared to be shared across samples. Hence, further investigation of these regions may indicate that they bear relevance as imaging biomarkers for ALS, particularly if they localize to known indicator regions such as the CST and precentral gyrus. Furthermore, the comparable localization of some feature clusters further underline the capabilities of features generated using V-CoHOG on imaging data which has not been registered or normalized.

5.2 Future Considerations

Additional work could be applied to a number of areas investigated in this thesis. Areas of interest visible in the feature segmentation maps may warrant further investigation as they could correspond to imaging biomarkers that could aid in patient diagnosis. Beyond ALS, there remain additional AD datasets apart from ADNI which could be considered, in addition to ALS mimics such as multiple Sclerosis. Furthermore, the suggested method is not restricted to brain images and could therefore be applied to conditions which impact various areas of the body. Because the nature of the generated features is not strictly suited to describing anatomical properties, but rather properties of image textures, our method could even be applied to non-medical classification tasks where volume data is available.

There are also a number of ways in which our method could be refined, including:

- Streamlining the parameter search process: Determining the optimal parameters for V-CoHOG is one of the least well-defined and potentially most prolonged

steps of the method. Rather than use a grid search on a single parameter basis, a more sophisticated approach, such as a Bayesian or gradient method, could allow for the determination of superior local maxima in less time.

- Extending ROI selection capabilities: As the ROI included for classification increases in size, classification performance may both increase or decrease. This could be due to the inclusion of additional biomarker regions, which would facilitate increased discrimination capabilities, or due to increases in noise, which can mislead the classifier model. Using an ROI cropping process which is more relative could help ensure that the ROIs between subjects are more comparable, even when there are differences in characteristics such as volume size. Via the use of more intricate segmentation masks, it may even be possible to crop the ROI based on neuroanatomical boundaries, as opposed to a cuboid selection.

- Using a combination of features: The feature vector generated by V-CoHOG is a concatenation of normalized CMs. Additional CMs can be calculated using a variety of GO types and CM calculation planes and then concatenated to form a larger feature vector with potential for greater classification capability. Nevertheless, this approach could also increase the amount of feature noise which could reduce the method's performance. Hence, concatenation could occur after feature selection with only the best features selected for different parameter combinations being added to the final feature vector.

Bibliography

[1] R. M. Haralick, K. S. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. 3, pp. 610–621, 1973.

[2] J. P. Folador, M. C. S. Santos, L. M. D. Luiz, L. A. P. S. de Souza, M. F. Vieira, A. A. Pereira, and A. de Oliveira Andrade, "On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people with parkinson's disease," *Medical & Biological Engineering & Computing*, vol. 59, pp. 195–214, 2021.

[3] J. Zhang and T. Tan, "Brief review of invariant texture analysis methods," *Pattern Recognit.*, vol. 35, pp. 735–747, 2002.

[4] S. Matthies, H. R. Wenk, and G. W. Vinel, "Some basic concepts of texture analysis and comparison of three methods to calculate orientation distributions from pole figures," *Journal of Applied Crystallography*, vol. 21, pp. 285–304, 1988.

[5] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 2 4, pp. 429–41, 1993.

[6] A. Materka and M. Strzelecki, "Texture analysis methods - a review," 1998.

[7] D. Wang, V. Haese-Coat, A. Bruno, and J. Ronsin, "Texture classification and segmentation based on iterative morphological decomposition," *J. Vis. Commun. Image Represent.*, vol. 4, pp. 197–214, 1993.

[8] G. M. M. E. Elahi, S. Kalra, L. Zinman, A. Genge, L. Korngut, and Y.-H. Yang, "Texture classification of mr images of the brain in als using m-cohog: A multi-center study," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 79, p. 101 659, 2020.

[9] S. Livens, P. Scheunders, G. V. de Wouwer, and D. V. Dyck, "Wavelets for texture analysis, an overview," 1997.

[10] G LoweDavid, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 2004.

[11] R. Porter and C. N. Canagarajah, "A robust automatic clustering scheme for image segmentation using wavelets," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 5 4, pp. 662–5, 1996.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 886–893 vol. 1, 2005.

[13] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," in *PSIVT*, 2009.

[14] G. Lian, "Pedestrian detection using quaternion histograms of oriented gradients," *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pp. 415–419, 2020.

[15] H. Bahri, M. Chouchene, F. Sayadi, and M. Atri, "Real-time moving human detection using hog and fourier descriptor based on cuda implementation," *Journal of Real-Time Image Processing*, pp. 1 –16, 2019.

[16] N. Andavarapu and V. K. Vatsavayi, "Weighted cohog (w-cohog) feature extraction for human detection," in *SocProS*, 2015.

[17] S. Iwata and S. Enokida, "Object detection based on multiresolution cohog," in *ISVC*, 2014.

[18] M. Zaffar, S. Ehsan, M. Milford, and K. D. Mcdonald-Maier, "Cohog: A lightweight, compute-efficient, and training-free visual place recognition technique for changing environments," *IEEE Robotics and Automation Letters*, vol. 5, pp. 1835–1842, 2020.

[19] N. E. Buch, J. Orwell, and S. A. Velastín, "3d extended histogram of oriented gradients (3dhog) for classification of road users in urban scenes," in *BMVC*, 2009.

[20] S. Cheng, A. Asthana, S. Zafeiriou, J. Shen, and M. Pantic, "Real-time generic face tracking in the wild with cuda," in *MMSys '14*, 2014.

[21] A. Parvez and A. Phadke, "Efficient implementation of glcm based texture feature computation using cuda platform," *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pp. 296–300, 2017.

[22] V. A. Prisacariu and I. D. Reid, "Fasthog – a real-time gpu implementation of hog," 2011.

[23] NVIDIA Corporation, *Cuda c++ programming guide*.

[24] Y. Nagamine, K. Yoshihiro, M. Shibata, H. Yamada, S. Enokida, and H. Tamukoh, "A hardware-oriented algorithm of gmm-mrcohog for high-performance human detection by an fpga," in *Other Conferences*, 2021.

[25] J. R. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with cuda," *2008 IEEE Hot Chips 20 Symposium (HCS)*, pp. 1–2, 2008.

[26] W. G. Bradley, "History of medical imaging.," *Proceedings of the American Philosophical Society*, vol. 152 3, pp. 349–61, 2008.

[27] B. M. W. Tsui, M. B. Smith, and K. K. Shung, "Principles of medical imaging," 1992.

[28] A. W. Barritt, M. C. Gabel, M. Cercignani, and P. N. Leigh, "Emerging magnetic resonance imaging techniques and analysis methods in amyotrophic lateral sclerosis," *Frontiers in Neurology*, vol. 9, 2018.

[29] P. Suetens, "Fundamentals of medical imaging," 2002.

[30] A. Gandhamal, S. N. Talbar, S. S. Gajre, A. F. M. Hani, and D. Kumar, "Local gray level s-curve transformation - a generalized contrast enhancement technique for medical images," *Computers in biology and medicine*, vol. 83, pp. 120–133, 2017.

[31] S. C. L. Deoni, S. C. R. Williams, P. Jezzard, J. Suckling, D. G. M. Murphy, and D. K. Jones, "Standardized structural magnetic resonance imaging in multicentre studies using quantitative t 1 and t 2 imaging at 1.5 t," *NeuroImage*, vol. 40, pp. 662–671, 2008.

[32] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. H. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," *ArXiv*, vol. abs/1807.10225, 2018.

[33] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. N. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. M. Wardlaw, and D. Rueckert, "Gan augmentation: Augmenting training data using generative adversarial networks," *ArXiv*, vol. abs/1810.10863, 2018.

[34] R. F. Cattell, S. Chen, and C. Huang, "Robustness of radiomic features in magnetic resonance imaging: Review and a phantom study," *Visual Computing for Industry, Biomedicine and Art*, vol. 2, 2019.

[35] L. Duron, D. Balvay, S. V. Perre, A. Bouchouicha, J. Savatovsky, J. Sadik, I. Thomassin-Naggara, L. S. Fournier, and A. Lecler, "Gray-level discretization impacts reproducible mri radiomics texture features," *PLoS ONE*, vol. 14, 2019.

[36] A. Chatterjee, M. Vallières, A. Dohan, I. R. Levesque, Y. Ueno, S. Saif, C. Reinhold, and J. Seuntjens, "Creating robust predictive radiomic models for data from independent institutions using normalization," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, pp. 210–215, 2019.

[37] S. Zarei, K. L. Carr, L. E. Reiley, K. Diaz, O. Guerra, P. Altamirano, W. Pagani, D. Lodin, G. Orozco, and Á. Chinea, "A comprehensive review of amyotrophic lateral sclerosis," *Surgical Neurology International*, vol. 6, 2015.

[38] W. Robberecht and T. E. J. Philips, "The changing scene of amyotrophic lateral sclerosis," *Nature Reviews Neuroscience*, vol. 14, pp. 248–264, 2013.

[39] K. Takei, K. Tsuda, F. Takahashi, M. Hirai, and J. M. Palumbo, "An assessment of treatment guidelines, clinical practices, demographics, and progression of disease among patients with amyotrophic lateral sclerosis in japan, the united states, and europe," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 18, pp. 88 –97, 2017.

[40] D. Harrison, P. Mehta, M. A. van Es, E. W. Stommel, V. Drory, B. Nefussy, L. H. van den Berg, J. Crayle, and R. S. Bedlack, "
als reversals": Demographics, disease characteristics, treatments, and co-morbidities," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 19, pp. 495 –499, 2018.

[41] P. E. Ludwig, S. C. Dulebohn, K. M. Metcalf, and S. Bhimji, "Neuroanatomy, central nervous system (cns)," 2017.

[42] P. Masrori and P. van Damme, "Amyotrophic lateral sclerosis: A clinical review," *European Journal of Neurology*, vol. 27, pp. 1918 –1929, 2020.

[43] S. Kalra, H.-P. Müller, A. Ishaque, L. Zinman, L. Korngut, A. Genge, C. Beaulieu, R. Frayne, S. J. Graham, and J. Kassubek, "A prospective harmonized multicenter dti study of cerebral white matter degeneration in als," *Neurology*, vol. 95, e943 –e952, 2020.

[44] T. Dhollander, "Tractography," Jun. 2018.

[45] N. Stifani, "Motor neurons and the generation of spinal motor neuron diversity," *Frontiers in Cellular Neuroscience*, vol. 8, 2014.

[46] R. Balendra and A. M. Isaacs, "C9orf72-mediated als and ftd: Multiple pathways to disease," *Nature Reviews Neurology*, vol. 14, pp. 544–558, 2018.

[47] S. Byrne, C. Walsh, C. Lynch, P. Bede, M. Elamin, K. P. Kenna, R. L. McLaughlin, and O. Hardiman, "Rate of familial amyotrophic lateral sclerosis: A systematic review and meta-analysis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 82, pp. 623 –627, 2010.

[48] S. Saxena and P. Caroni, "Selective neuronal vulnerability in neurodegenerative diseases: From stressor thresholds to degeneration," *Neuron*, vol. 71, pp. 35–48, 2011.

[49] J. M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. C. Hilt, B. Thurmond, and A. Nakanishi, "The alsfrs-r: A revised als functional rating scale that incorporates assessments of respiratory function," *Journal of the Neurological Sciences*, vol. 169, pp. 13–21, 1999.

[50] J. Rooney, T. Burke, A. Vajda, M. Heverin, and O. Hardiman, "What does the alsfrs-r really measure? a longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 88, pp. 381 –385, 2016.

[51] R. P. A. van Eijk, A. D. de Jongh, S. Nikolakopoulos, C. J. Mcdermott, M. J. C. Eijkemans, K. C. B. Roes, and L. H. van den Berg, "An old friend who has overstayed their welcome: The alsfrs-r total score as primary endpoint for als clinical trials," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 22, pp. 300 –307, 2021.

[52] O. M. Peters, M. Ghasemi, and R. H. Brown, "Emerging mechanisms of molecular pathology in als.," *The Journal of clinical investigation*, vol. 125 5, pp. 1767–79, 2015.

[53] X. Yang, Y. Ji, W. Wang, L. Zhang, Z. Chen, M. Yu, Y. Shen, F. Ding, X. Gu, and H. Sun, "Amyotrophic lateral sclerosis: Molecular mechanisms, biomarkers, and therapeutic strategies," *Antioxidants*, vol. 10, 2021.

[54] A. Jeromin and R. P. Bowser, "Biomarkers in neurodegenerative diseases.," *Advances in neurobiology*, vol. 15, pp. 491–528, 2017.

[55] U. Klickovic, L. Zampedri, C. D. J. Sinclair, S. J. Wastling, K. Trimmel, R. S. Howard, A. Malaspina, N. Sharma, K. C. L. Sidle, A. Emira, S. Shah, T. A. Yousry, M. G. Hanna, L. Greensmith, J. M. Morrow, J. S. Thornton, and P. Fratta, "Skeletal muscle mri differentiates sbma and als and correlates with disease severity," *Neurology*, vol. 93, e895 –e907, 2019.

[56] J. Labra, P. Menon, K. Byth, S. Morrison, and S. Vucic, "Rate of disease progression: A prognostic biomarker in als," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 87, pp. 628 –632, 2015.

[57] M. Dadar, A. L. Manera, L. Zinman, L. Korngut, L. Korngut, A. Genge, S. J. Graham, R. Frayne, R. Frayne, D. L. Collins, and S. Kalra, "Cerebral atrophy in amyotrophic lateral sclerosis parallels the pathological distribution of tdp43," *Brain Communications*, vol. 2, 2020.

[58] J. Jin, F. Hu, Q. Zhang, R. Jia, and J. xia Dang, "Hyperintensity of the corticospinal tract on flair: A simple and sensitive objective upper motor neuron degeneration marker in clinically verified amyotrophic lateral sclerosis," *Journal of the Neurological Sciences*, vol. 367, pp. 177–183, 2016.

[59] A. Ishaque, R. Maani, J. Satkunam, P. Seres, D. Mah, A. H. Wilman, S. Naik, Y.-H. Yang, and S. Kalra, "Texture analysis to detect cerebral degeneration in amyotrophic lateral sclerosis," *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, vol. 45, pp. 533 –539, 2018.

[60] A. Ishaque, D. Mah, P. Seres, C. C. Luk, W. Johnston, S. Chenji, C. Beaulieu, Y.-H. Yang, and S. Kalra, "Corticospinal tract degeneration in als unmasked in t1-weighted images using texture analysis," *Human Brain Mapping*, vol. 40, pp. 1174 –1183, 2019.

[61] E. Kasper, C. Schuster, J. Machts, J. Kaufmann, D. Bittner, S. Vielhaber, R. Benecke, S. J. Teipel, and J. Prudlo, "Microstructural white matter changes underlying cognitive and behavioural impairment in als – an in vivo study using dti," *PLoS ONE*, vol. 9, 2014.

[62] F. Trojsi, F. di Nardo, M. Siciliano, G. Caiazzo, C. Femiano, C. Passaniti, D. Ricciardi, A. Russo, A. Bisecco, S. Esposito, M. R. Monsurrò, M. Cirillo, G. Santangelo, F. Esposito, and G. Tedeschi, "Frontotemporal degeneration in amyotrophic lateral sclerosis (als): A longitudinal mri one-year study," *CNS Spectrums*, vol. 26, pp. 258 –267, 2020.

[63] H.-P. Müller, M. R. Turner, J. Grosskreutz, S. Abrahams, P. Bede, V. Govind, J. Prudlo, A. C. Ludolph, M. Filippi, and J. Kassubek, "A large-scale multicentre cerebral diffusion tensor imaging study in amyotrophic lateral sclerosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 87, pp. 570 –579, 2016.

[64] S. Kalra, M. U. Khan, L. Barlow, C. Beaulieu, M. Benatar, H. R. Briemberg, S. Chenji, M. G. Clua, A. Cormier, S. Das, A. Dionne, N. Dupré, D. Emery, D. T. Eurich, R. Frayne, A. Genge, S. Gibson, S. J. Graham, C. C. Hanstock, A. Ishaque, J. Joseph, J. Keith, L. Korngut, D. B. Krebs, C. R. McCreary, F. Pattany, P. Seres, C. L. Shoesmith, T. Szekeres, F. Tam, R. C. Welsh, A. H. Wilman, Y. Yang, Y. R. Yunusova, L. Zinman, and C. A. N. Consortium, "The canadian als neuroimaging consortium (calsnic) - a multicentre platform for standardized imaging and clinical studies in als," *medRxiv*, 2020.

[65] R. Maani, Y.-H. Yang, and S. Kalra, "Voxel-based texture analysis of the brain," *PLoS ONE*, vol. 10, 2015.

[66] D. Ta, M. R. H. Khan, A. Ishaque, P. Seres, D. T. Eurich, Y.-H. Yang, and S. Kalra, "Reliability of 3d texture analysis: A multicenter mri study of the brain," *Journal of Magnetic Resonance Imaging*, vol. 51, 2019.

[67] MATLAB, *9.11.0.1873467 (R2021b)*. Natick, Massachusetts: The MathWorks Inc., 2022.

[68] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, 2002.

[69] A. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1200–1205, 2015.

[70] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, pp. 16–28, 2014.

[71] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *ML*, 1992.

[72] I. Kononenko, E. Simec, and M. Robnik-Sikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, pp. 39–55, 2004.

[73] R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of biomedical informatics*, vol. 85, pp. 189–203, 2018.

[74] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 2004.

[75] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of svm- and elm-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, 2014.

[76] E. Fix and J. L. Hodges, "Discriminatory analysis - nonparametric discrimination: Consistency properties," *International Statistical Review*, vol. 57, p. 238, 1989.

[77] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, pp. 193–202, 2004.

[78] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *NIPS*, 2007.

[79] J. S. Cramer, "The origins of logistic regression," *Econometrics eJournal*, 2002.

[80] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, pp. 21–45, 2006.

[81] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, 2017.

[82] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017.

[83] A. Ishaque, D. Ta, M. R. H. Khan, L. Zinman, L. Korngut, A. Genge, A. Dionne, H. R. Briemberg, C. C. Luk, Y.-H. Yang, C. Beaulieu, D. J. Emery, D. T. Eurich, R. Frayne, S. Graham, A. H. Wilman, N. Dupré, and S. Kalra, "Distinct patterns of progressive gray and white matter degeneration in amyotrophic lateral sclerosis," *Human Brain Mapping*, vol. 43, pp. 1519 –1534, 2021.

[84] B. R. Brooks, R. G. Miller, M. Swash, and T. L. Munsat, "El escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, vol. 1, pp. 293 –299, 2000.

[85] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[86] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[87] N. Ma, X. Zhang, H. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," *ArXiv*, vol. abs/1807.11164, 2018.

[88] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[89] J. Weller and A. E. Budson, "Current understanding of alzheimer's disease diagnosis and treatment," *F1000Research*, vol. 7, 2018.

[90] J. Park, "Mortality from alzheimer's disease in canada: A multiple-cause-of-death analysis, 2004 to 2011.," *Health reports*, vol. 27 5, pp. 17–21, 2016.

[91] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, "Alzheimer's disease neuroimaging initiative (adni)," *Neurology*, vol. 74, pp. 201 –209, 2010.